

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 17, 2004, 09:44:37 ; Search time 41 Seconds  
(without alignments)  
677.491 Million cell updates/sec

Title: US-09-589-777C-2\_COPY\_1\_175  
Perfect score: 923  
Sequence: 1 HTHQDFQPVHLVALNTPLS.....RLLEQKAASCHNSYIVLCIE 175

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A Geneseq 19Jun03:\*  
1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*  
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*  
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*  
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*  
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	923	100.0	184	22 AAB49380	Murine endostatin
2	923	100.0	191	21 AAB28398	Murine endostatin.
3	923	100.0	191	23 AAU77950	Amino acid sequenc
4	923	100.0	207	23 ABB79902	Mouse endostatin.
5	920	99.7	183	23 AAM49504	Mouse endostatin p
6	920	99.7	184	20 AAU08689	Murine endostatin
7	920	99.7	184	21 AAY70258	Murine angiogenesi
8	920	99.7	184	23 ABG31793	Human endostatin p
9	920	99.7	207	22 AAE02031	Murine endostatin

10	920	99.7	207	22 AAB71930	Murine endostatin
11	920	99.7	218	20 AAY08691	Murine gene therap
12	920	99.7	580	20 AAY08692	Murine gene therap
13	920	99.7	684	20 AAY25114	Mouse alpha1 (XVII
14	920	99.7	1288	18 AAW26328	Mouse alpha-1 coll
15	920	99.7	1288	20 AAW92297	Endostatin protein
16	918	99.5	184	20 AAY18409	Anti-angiogenic en
17	874	94.7	185	20 AAY06197	Human Endostatin(T
18	813	88.1	180	22 AAU00899	Human Endostatin(T
19	813	88.1	181	22 AAU00898	Human endostatin.
20	813	88.1	182	21 AAB28399	Human endostatin p
21	813	88.1	182	21 AAY94323	Human endostatin p
22	813	88.1	182	21 AAY59622	Human endostatin(T
23	813	88.1	182	22 AAU00897	Human Endostatin(T
24	813	88.1	182	23 AAU77951	Amino acid sequenc
25	813	88.1	183	20 AAY08693	Human endostatin p
26	813	88.1	183	20 AAY02113	SEQ ID 76 of WO991
27	813	88.1	183	21 AAB30493	Amino acid sequenc
28	813	88.1	183	21 AAB16451	Human endostatin p
29	813	88.1	183	21 AAY90771	Human angiogenesis
30	813	88.1	183	21 AAY70252	Human angiogenesis
31	813	88.1	183	22 AAU00896	Human Endostatin(T
32	813	88.1	183	22 AAB49379	Human Endostatin S
33	813	88.1	183	23 ABB79901	Human endostatin p
34	813	88.1	183	23 AAU97132	Human endostatin.
35	813	88.1	183	23 AAM49503	Human endostatin p
36	813	88.1	183	23 AAM48895	Human endostatin p
37	813	88.1	183	24 AAG79753	Human endostatin.
38	813	88.1	184	23 ABG31794	Murine endostatin
39	813	88.1	193	21 AAW90877	Human HMW endostat
40	813	88.1	195	21 AAW90874	Human HMW endostat
41	813	88.1	216	21 AAB30495	Amino acid sequenc
42	813	88.1	275	23 AAU76689	Synthetic plasmid
43	813	88.1	310	23 AAU76688	Human collagen XVI
44	813	88.1	513	23 ABG73586	Human Endostatin/I
45	813	88.1	682	23 ABP41878	Human ovarian anti

ALIGNMENTS

RESULT 1  
AAB49380  
ID AAB49380 standard; Protein; 184 AA.  
XX  
AC AAB49380;  
XX  
DT 02-MAR-2001 (first entry)  
XX  
DE Murine endostatin SEQ ID NO: 4.  
XX  
KW Endostatin; antiangiogenic; angiogenesis; human; mouse; chicken;  
KW cancer; inflammation; angiogenesis-dependent disease.  
XX  
OS Mus musculus.  
XX  
PN WO200067771-A1.  
XX  
PD 16-NOV-2000.  
XX  
PF 02-MAY-2000; 2000WO-US12063.  
XX  
PR 06-MAY-1999; 99US-0132907.  
PR 14-JUL-1999; 99US-0353333.  
XX  
PA (BURN-) BURNHAM INST.  
XX  
PI Vuori K;  
XX  
DR WPI; 2001-040937/05.  
DR N-PSDB; AAC88290.  
XX  
PT Endostatin peptide comprising at least four endostatin amino acid

PT residues are e.g. angiogenesis inhibitors for treating cancer and  
PT diabetic retinopathy -  
XX Disclosure; Fig 1; 146pp; English.  
XX The present invention provides endostatin peptides which can be used in  
CC the modulation of angiogenesis. This is useful in the treatment of  
CC cancers, inflammation, rheumatoid arthritis, chronic articular  
CC rheumatism, psoriasis, disorders associated with inopportune invasion of  
CC vessels such as diabetic retinopathy, neovascular glaucoma, retinopathy  
CC of prematurity, macular degeneration, corneal graft rejection,  
CC retrolental fibroplasia, rubeosis, capillary proliferation in  
CC atherosclerotic plaques, and osteoporosis. Other angiogenesis-dependent  
CC diseases include Osler-Webber syndrome, myocardial angiogenesis, plaque  
CC neovascularisation, telangiectasia, haemophiliac joints and wound  
CC granulation. In addition, the peptides can be used as birth control  
CC agents.  
XX  
SQ Sequence 184 AA;  
Query Match 100.0%; Score 923; DB 22; Length 184;  
Best Local Similarity 100.0%; Pred. No. 2.8e-102;  
Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 HTHQDFQPVHLVALNTPLSGMGRGIRGADFQCFOQARAVGLSGTFRFLSSRLQDLYSI 60  
Db 1 HTHQDFQPVHLVALNTPLSGMGRGIRGADFQCFOQARAVGLSGTFRFLSSRLQDLYSI 60  
QY 61 VRRADRGSVPIVNLKDEVLSFSGSQQLQPGARIFSFDRDLRHPAWPQKSVW 120  
Db 61 VRRADRGSVPIVNLKDEVLSFSGSQQLQPGARIFSFDRDLRHPAWPQKSVW 120  
QY 121 HGSDPSGRRRLMESYCETWRTTGTGATGQASSLLSGLLEQKAASCHNSYIVLCIE 175  
Db 121 HGSDPSGRRRLMESYCETWRTTGTGATGQASSLLSGLLEQKAASCHNSYIVLCIE 175  
RESULT 2  
AAB28398  
ID AAB28398 standard; Protein; 191 AA.  
XX  
AC AAB28398;  
XX  
DT 19-FEB-2001 (first entry)  
XX Murine endostatin.  
XX Murine; endostatin; cytostatic; antiproliferative;  
KW vascular endothelial growth factor; VEGF; antibody; VEGF2 receptor;  
KW cancer; vascularised solid tumour.  
XX  
OS Mus sp.  
XX WO200064946-A2.  
XX  
PD 02-NOV-2000.  
XX  
PF 28-APR-2000; 2000WO-US11367.  
XX  
PR 28-APR-1999; 99US-0131432.  
XX (TEXA ) UNIV TEXAS SYSTEM.  
XX Thorpe PE, Brekken RA;  
XX WPI; 2000-687317/67.  
DR N-PSDB; AAC67777.  
XX  
XX Immunogenic composition for the treatment and diagnosis of cancer  
PT comprises an anti-VEGF (vascular endothelial growth factor) antibody  
PT binding the same epitope as the monoclonal antibody ATCC PTA 1595 -  
XX  
PS Example 10; Page 290-291; 298pp; English.

XX The present invention relates to anti-Vascular Endothelial Growth Factor  
CC (VEGF) antibodies that bind to the same epitope as the monoclonal  
CC antibody ATCC PTA 1595 and which significantly inhibit VEGF binding to  
CC the VEGF receptor VEGFR2, without inhibiting VEGF binding to the VEGF  
CC receptor VEGFR1. The present sequence is murine endostatin. Endostatin  
CC may be conjugated onto the anti-VEGF antibodies of the present invention.  
CC The anti-VEGF antibodies of the present invention are useful for the  
CC treatment and diagnosis of cancer, especially vascularised solid tumours.  
XX  
SQ Sequence 191 AA;  
Query Match 100.0%; Score 923; DB 21; Length 191;  
Best Local Similarity 100.0%; Pred. No. 2.9e-102;  
Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 HTHQDFQPVHLVALNTPLSGMGRGIRGADFQCFOQARAVGLSGTFRFLSSRLQDLYSI 60  
Db 8 HTHQDFQPVHLVALNTPLSGMGRGIRGADFQCFOQARAVGLSGTFRFLSSRLQDLYSI 67  
QY 61 VRRADRGSVPIVNLKDEVLSFSGSQQLQPGARIFSFDRDLRHPAWPQKSVW 120  
Db 68 VRRADRGSVPIVNLKDEVLSFSGSQQLQPGARIFSFDRDLRHPAWPQKSVW 127  
QY 121 HGSDPSGRRRLMESYCETWRTTGTGATGQASSLLSGLLEQKAASCHNSYIVLCIE 175  
Db 128 HGSDPSGRRRLMESYCETWRTTGTGATGQASSLLSGLLEQKAASCHNSYIVLCIE 182  
RESULT 3  
AAU77950  
ID AAU77950 standard; Protein; 191 AA.  
XX  
AC AAU77950;  
XX  
DT 02-JUL-2002 (first entry)  
XX  
DE Amino acid sequence for mouse endostatin.  
XX  
KW Mouse; immunoconjugate; anti-vascular endothelial growth factor antibody;  
KW anti-VEGF antibody; monoclonal antibody 2C3 ATCC PTA 1595; VEGF receptor;  
KW VEGFR2; KDR/Flk-1; VEGFR1; Flt-1; angiogenesis; macular degeneration;  
KW ocular neovascular disease; cancer; vascularised solid tumour; AIDS;  
KW metastatic tumour; endothelial cell proliferation; inflammatory disorder;  
KW atherosclerosis; diabetic retinopathy; corneal graft rejection;  
KW acquired immune deficiency syndrome; infection; restenosis; fungal ulcer;  
KW sickle cell anaemia; endometriosis; endostatin.  
XX  
OS Mus sp.  
XX  
XX AU200179401-A.  
XX  
XX 06-DEC-2001.  
XX  
PF 12-OCT-2001; 2001AU-0079401.  
XX  
PR 28-APR-2000; 2000AU-0048049.  
PR 12-OCT-2001; 2001AU-0079401.  
XX  
PA (TEXA ) UNIV TEXAS SYSTEM.  
XX  
PI Thorpe PE, Brekken RA;  
XX  
XX WPI; 2002-281368/33.  
DR N-PSDB; ABK47719.  
XX  
XX Immunoconjugate compositions for treating cancer by inhibiting  
PT angiogenesis and for delivering a diagnostic agent to tumour, comprises  
PT anti-vascular endothelial growth factor antibody attached to a  
PT biological agent -  
XX  
PS Example 10; Page 11-12 (Sequence listing); 300pp; English.

CC The present invention relates to antibody-based compositions comprising  
CC an immunconjugate such as anti-vascular endothelial growth factor  
CC (VEGF) antibody (Ab) (or its antigen-binding fragment), attached to a  
CC biological agent, where the Ab binds to the same epitope as the  
CC monoclonal antibody (Mab) 2C3 ATCC PTA 1595, and significantly inhibits  
CC VEGF binding to the VEGF receptor VEGFR2 (KDR/Flk-1) without inhibiting  
CC VEGF binding to the VEGF receptor VEGFR1 (Flt-1). The compositions  
CC of the invention are useful in therapy, and diagnosis, for inhibiting  
CC angiogenesis in an animal having ocular neovascular disease or macular  
CC degeneration, and for delivering a biological agent to a vascularised  
CC tumour. The compositions can also be used for treating cancer and  
CC subjects at risk of developing, a vascularised solid tumour, a metastatic  
CC tumour or metastases from a primary tumour. The composition is useful  
CC for specifically inhibiting VEGF-induced endothelial cell proliferation,  
CC without significantly inhibiting VEGF-induced macrophage, osteoclast or  
CC chondroclast function. The compositions can be used for treating various  
CC diseases such as inflammatory disorders, atherosclerosis, diabetic  
CC retinopathy, restenosis, acquired immune deficiency syndrome (AIDS),  
CC blood borne tumours, corneal graft rejection, Crohn's disease, fungal  
CC ulcers, infections, sickle cell anaemia, and endometriosis. The present  
CC sequence represents mouse endostatin. Endostatin may be attached or  
CC functionally associated with anti-VEGF antibodies.

XX Sequence 191 AA;

Query Match 100.0%; Score 923; DB 23; Length 191;  
Best Local Similarity 100.0%; Pred. No. 2.9e-102;  
Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HTHQDFQPVHLVALNTPLSGMGRGIRGADFQCQQAQAVGLSGTFFRAFLSSRLQDLYSI 60  
DB 8 HTHQDFQPVHLVALNTPLSGMGRGIRGADFQCQQAQAVGLSGTFFRAFLSSRLQDLYSI 67  
QY 61 VRRADRGSPVIVNLKDEVLSPSWDSLSFGSQGQLQPGARIFSFDRDVLRRHPWPQKSVW 120  
DB 68 VRRADRGSPVIVNLKDEVLSPSWDSLSFGSQGQLQPGARIFSFDRDVLRRHPWPQKSVW 127  
QY 121 HGSDPSGRRLMESYCETWRTTGTGATGQASSLLSRLLEQKAASCHNSYIVLCIE 175  
DB 128 HGSDPSGRRLMESYCETWRTTGTGATGQASSLLSRLLEQKAASCHNSYIVLCIE 182

RESULT 4  
ABB79902  
ID ABB79902 standard; Protein; 207 AA.

XX ABB79902;  
XX 05-DEC-2002 (first entry)  
XX Mouse endostatin.  
XX Endostatin; mouse; ophthalmological; ocular neovascularisation;  
KW choroidal neovascularisation; gene therapy.

OS Mus musculus.  
FH Key Location/Qualifiers  
FT Misc-difference 117 /note= "encoded by GTG"  
PN WO200267971-A2.

XX 06-SEP-2002.  
XX 21-FEB-2002; 2002WO-US05336.  
XX 22-FEB-2001; 2001US-270787P.  
PR 04-APR-2001; 2001US-281296P.

XX (NOVS ) NOVARTIS AG.  
XX Brazzell RK, Campochiaro PA, Dixon KH;

XX WPI; 2002-698636/75.  
DR N-PSDB; ABQ81194.  
XX Treating or preventing choroidal neovascularization comprises  
PT increasing the amount of endostatin in ocular tissues of afflicted  
PT individuals to a choroidal neovascularization inhibiting level -  
XX Disclosure; Page 40; 44pp; English.

XX The present sequence is the protein sequence of murine endostatin  
CC plus the murine Ig kappa leader sequence. A claimed method for the  
CC treatment of ocular neovascularisation, especially choroidal  
CC neovascularisation, involves increasing the level of endostatin in  
CC ocular tissue, especially the level of human endostatin (see  
CC ABB79901), its fragment, derivative or variant. The increase is  
CC effected by administering a viral vector, especially an adenovirus,  
CC adeno-associated virus, a retrovirus or lentivirus vector,  
CC comprising an endostatin-encoding nucleic acid. Cells secreting  
CC endostatin may be encapsulated and implanted within an individual.  
CC The method is used when ocular neovascularisation is caused by  
CC histoplasmosis, pathological myopia, angioid streaks, anterior  
CC ischaemic optic neuropathy, bacterial endocarditis, Best's disease,  
CC birdshot retinochoroidopathy, choroidal haemangioma, choroidal  
CC naevi, choroidal nonperfusion, choroidal osteomas, choroidal  
CC rupture, chorioideraemia, chronic retinal detachment, coloboma of  
CC the retina, Drusen, endogenous Candida endophthalmitis,  
CC extrapapillary hamartoma of the retinal pigmented epithelium,  
CC fundus flavimaculatus, idiopathic, macular hole, malignant  
CC melanoma, membranoproliferative glomerulonephritis (type II),  
CC metallic intraocular foreign body, morning glory disc syndrome,  
CC multiple evanescent white-dot syndrome, neovascularisation of ora  
CC serrata, operating microscope burn, optic nerve head pits,  
CC photocoagulation, punctate inner choroidopathy, rubella,  
CC sarcoidosis, seriginous or geographic choroiditis, subretinal  
CC fluid drainage, tiled disc syndrome, Toxoplasma retinochoroiditis,  
CC tuberculosi, Vogt-Koyanagi-Harada syndrome, diabetic retinopathy,  
CC non-diabetic retinopathy, brain vein occlusion, central retinal  
CC vein occlusion, retinopathy in premature infants, rubeosis iridis,  
CC neovascular glaucoma, perifoveal telangiectasis, sickle cell  
CC retinopathy, Eale's disease, retinal vasculitis, Von Hippel  
CC Lindau disease, radiation retinopathy, retinal cryoinjury,  
CC retinitis pigmentosa, retinochoroidal coloboma, corneal  
CC neovascularisation due to herpes simplex keratitis, corneal ulcers,  
CC keratoplasty, pterygia and trauma (all claimed).

XX Sequence 207 AA;

Query Match 100.0%; Score 923; DB 23; Length 207;  
Best Local Similarity 100.0%; Pred. No. 3.3e-102;  
Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HTHQDFQPVHLVALNTPLSGMGRGIRGADFQCQQAQAVGLSGTFFRAFLSSRLQDLYSI 60  
DB 24 HTHQDFQPVHLVALNTPLSGMGRGIRGADFQCQQAQAVGLSGTFFRAFLSSRLQDLYSI 83  
QY 61 VRRADRGSPVIVNLKDEVLSPSWDSLSFGSQGQLQPGARIFSFDRDVLRRHPWPQKSVW 120  
DB 84 VRRADRGSPVIVNLKDEVLSPSWDSLSFGSQGQLQPGARIFSFDRDVLRRHPWPQKSVW 143  
QY 121 HGSDPSGRRLMESYCETWRTTGTGATGQASSLLSRLLEQKAASCHNSYIVLCIE 175  
DB 144 HGSDPSGRRLMESYCETWRTTGTGATGQASSLLSRLLEQKAASCHNSYIVLCIE 198

RESULT 5  
AAM49504  
ID AAM49504 standard; Protein; 183 AA.

XX AAM49504;  
XX 07-MAY-2002 (first entry)



XX PD 02-MAR-2000.  
XX PF 25-AUG-1999; 99WO-US19329.  
XX PR 25-AUG-1998; 98US-0097883.  
XX PA (LEXI-) LEXINGEN PHARM CORP.  
XX PI Lo K, Li Y, Gillies SD;  
XX WPI; 2000-237616/20.  
DR N-PSDB; AA251299.  
XX Novel fusion protein of angiotensin or endostatin and an immunoglobulin  
PT FC region, useful for treating conditions mediated by angiogenesis,  
PT such as rheumatoid arthritis, tumors and macular degeneration -  
XX Example 5; Pages 48-49; 68pp; English.  
XX The patent discloses a DNA molecule encoding a fusion protein comprising  
CC a signal sequence, an immunoglobulin Fc region, and an angiogenesis  
CC inhibitor selected from angiotensin, endostatin, a plasminogen fragment  
CC having angiotensin activity, a collagen XVIII fragment having endostatin  
CC activity, or combinations of them. The fusion protein (immunofusin) is  
CC used to inhibit angiogenesis and to treat diseases or conditions mediated  
CC by angiogenesis. Conditions that may be treated include solid tumors,  
CC blood born tumors, tumor metastasis, benign tumors including  
CC haemangiomas, acoustic neuromas, neurofibromas, trachomas and pyrogenic  
CC granulomas, rheumatoid arthritis, psoriasis, ocular angiogenic diseases  
CC e.g. diabetic retinopathy, retinopathy of prematurity, macular  
CC degeneration, corneal graft rejection, neovascular glaucoma, retrolental  
CC fibroplasia, rubeosis and Osler-Webber syndrome; myocardial angiogenesis,  
CC plaque neovascularisation, telangiectasia, haemophilic joints,  
CC angiofibroma, wound granulation, and excessive or abnormal stimulation of  
CC endothelial cells, intestinal cells, atherosclerosis, scleroderma and  
CC hypertrophic scars, i.e. keloid scars. The DNA constructs may be used  
CC in gene therapy. The present sequence is a murine  
CC endostatin used in the construction of immunofusin containing murine  
CC immunoglobulin Fc fragment.  
XX Sequence 184 AA;  
SQ  
Query Match 99.7%; Score 920; DB 21; Length 184;  
Best Local Similarity 99.4%; Pred. No. 6.4e-102;  
Matches 174; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 HTHQDFQVHLVALNTPLSGMGRGIRGADFQCFCQARAVGLSGTFRAPLSSRLQDLYSI 60  
Db 1 HTHQDFQVHLVALNTPLSGMGRGIRGADFQCFCQARAVGLSGTFRAPLSSRLQDLYSI 60  
QY 61 VRRADRGSPVIVNLKDEVLSFSGSQGLQPGARIFSFDRDVLRHAPWPKSVW 120  
Db 61 VRRADRGSPVIVNLKDEVLSFSGSQGLQPGARIFSFDRDVLRHAPWPKSVW 120  
QY 121 HGSDPSGRRLMESYCETWRTTGTATGQASSLLSRLLEQKAASCHNSYIVLCIE 175  
Db 121 HGSDPSGRRLMESYCETWRTTGTATGQASSLLSRLLEQKAASCHNSYIVLCIE 175  
RESULT 8  
ABG31793  
ID ABG31793 standard; Protein; 184 AA.  
XX  
AC ABG31793;  
XX  
DT 05-DEC-2002 (first entry)  
XX  
DE Human endostatin polypeptide.  
KW Human; endostatin; tumor; cancer; metastasis; cytostatic;  
KW antiangiogenic.  
XX

OS Homo sapiens.  
XX WO200268457-A2.  
PN  
XX 06-SEP-2002.  
PD  
XX 27-FEB-2002; 2002WO-IT00119.  
PF  
XX 27-FEB-2001; 2001IT-MI00394.  
XX  
XX (UTMI-) UNIV MILANO.  
PA  
XX Chillemi F, Vicentinie LMT, Francescato P;  
PI  
XX WPI; 2002-698655/75.  
DR  
XX New peptide useful for the preparation of medicaments with  
XX antiangiogenic activity that may be used in treating tumors or  
PT metastases, comprises a sequence corresponding to fragments of human  
PT endostatin -  
PT  
XX Disclosure; Fig 1; 24pp; English.  
PS  
XX The invention relates to peptide comprising 20-50 amino acids with  
XX sequences corresponding to the human endostatin polypeptide sequence, its  
CC salt or non-toxic derivative. The peptides are useful in the preparation  
CC of medicaments with antiangiogenic activity which may be useful in  
CC treating tumors or metastases. This sequence represents a human  
CC endostatin polypeptide.  
XX Sequence 184 AA;  
SQ  
Query Match 99.7%; Score 920; DB 23; Length 184;  
Best Local Similarity 99.4%; Pred. No. 6.4e-102;  
Matches 174; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 HTHQDFQVHLVALNTPLSGMGRGIRGADFQCFCQARAVGLSGTFRAPLSSRLQDLYSI 60  
Db 1 HTHQDFQVHLVALNTPLSGMGRGIRGADFQCFCQARAVGLSGTFRAPLSSRLQDLYSI 60  
QY 61 VRRADRGSPVIVNLKDEVLSFSGSQGLQPGARIFSFDRDVLRHAPWPKSVW 120  
Db 61 VRRADRGSPVIVNLKDEVLSFSGSQGLQPGARIFSFDRDVLRHAPWPKSVW 120  
QY 121 HGSDPSGRRLMESYCETWRTTGTATGQASSLLSRLLEQKAASCHNSYIVLCIE 175  
Db 121 HGSDPSGRRLMESYCETWRTTGTATGQASSLLSRLLEQKAASCHNSYIVLCIE 175  
RESULT 9  
AAE02031  
ID AAE02031 standard; Protein; 207 AA.  
XX  
AC AAE02031;  
XX  
DT 31-JUL-2001 (first entry)  
XX  
DE Murine endostatin fused to N-terminal secretion signal.  
XX  
KW Murine; endostatin; fusion protein; nucleotide-binding domain; NBD;  
KW ligand-binding domain; LBD; transcription regulating domain; TRD; cancer;  
KW zinc finger protein; ZFP; ligand-activated transcriptional regulator;  
KW gene regulation; gene therapy; cell proliferative disorder; psoriasis;  
KW pemphigus vulgaris; Behcet's syndrome; lipid histiocytosis.  
XX  
OS Mus sp.  
XX  
XX WO200130843-A1.  
PN  
XX 03-MAY-2001.  
PD  
XX 23-OCT-2000; 2000WO-EP10430.  
PF  
XX



PR 25-OCT-1999; 99US-0433042.  
PR 02-JUN-2000; 2000US-0586625.  
XX  
PA (NOVS ) NOVARTIS AG.  
PA (SCRI ) SCRIPPS RES INST.  
XX  
PI Barbas CF, Kadan M, Beerli R;  
XX  
XX WPI; 2001-308618/32.  
DR N-PSDB; AAD06108.  
XX

PT New fusion protein containing nucleotide-binding and ligand-binding  
PT domains, useful e.g. in gene therapy of cancer, provides  
PT ligand-activated control of gene expression -  
XX  
PS Example 19; Page 209; 218pp; English.

XX The invention relates to fusion protein comprising a nucleotide-binding  
CC domain (NBD), a ligand-binding domain (LBD) of an intracellular receptor  
CC (ICR) and a transcription regulating domain (TRD). NBD is a polydactyl  
CC zinc finger protein (ZFP), or a modular part of it, that interacts  
CC specifically with a contiguous sequence of at least 3 nucleotides. The  
CC fusion protein functions as a ligand-activated transcriptional regulator.  
CC The fusion protein and the nucleic acid encoding it, are used to regulate  
CC gene expression, particularly in gene therapy for treating malignant  
CC cell proliferative diseases (e.g. colon cancer, prostate cancer,  
CC renal-cell carcinoma) and non-malignant cell proliferative  
CC diseases (e.g. psoriasis, pemphigus vulgaris, Behcet's syndrome and  
CC lipid histiocytosis). The fusion protein and its DNA are also useful for  
CC treating diseases caused by viruses in humans/plants, genetic and/or  
CC acquired diseases. The fusion protein can be designed to target any  
CC selected gene (endogenous or exogenous), and can be made to have  
CC different selectivity or specificity for endogenous or exogenous ligands.  
CC The present sequence is murine endostatin fused to an N-terminal  
CC secretion signal. The corresponding cDNA sequence was used in the  
CC construction of left end shuttle plasmids containing regulatable  
CC transgene cassettes for evaluation of Cys2-His2 Zinc finger DNA binding  
CC domain (DBD)-Oestrogen receptor (ER) LBD regulators.

XX Sequence 207 AA;

Query Match 99.7%; Score 920; DB 22; Length 207;  
Best Local Similarity 99.4%; Pred. No. 7.6e-102;  
Matches 174; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 HTHQDFQPVHLVALNTPLSGMVGIRGADFQCFQQAQAVGLSGTFRFLSSRLQDLYSI 60  
DB 24 HTHQDFQPVHLVALNTPLSGMVGIRGADFQCFQQAQAVGLSGTFRFLSSRLQDLYSI 83  
QY 61 VRRADRGSPVIVNLKDEVLSPSWDSLFSGSQGLQPGARIFSFDRDLRHPAPQKSVW 120  
DB 84 VRRADRGSPVIVNLKDEVLSPSWDSLFSGSQGLQPGARIFSFDRDLRHPAPQKSVW 143  
QY 121 HGSDPSGRRLMESYCETWRTTGTATGQASSLLSGLLEQKAASCHNSYIVLCIE 175  
DB 144 HGSDPSGRRLMESYCETWRTTGTATGQASSLLSGLLEQKAASCHNSYIVLCIE 198

RESULT 10  
AAB71930  
ID AAB71930 standard; Protein; 207 AA.

XX AAB71930;

DT 10-MAY-2001 (first entry)

XX Murine endostatin attached to Ig-kappa leader sequence.

DE Mouse; endostatin; antitumour; cytostatic; antiarthritic; antipsoriatic;  
XX antidiabetic; ophthalmological; gene therapy; angiogenic inhibitor;  
KW adenoviral vector; diabetic retinopathy; cardiovascular disease;  
KW arthritis; psoriasis; cerebral oedema; intravascular coagulopathy;  
KW lymphoma; leukaemia; immunoglobulin; Ig; Ig-kappa.

XX Mus sp.  
OS WO200112830-A1.  
XX  
PN 22-FEB-2001.  
XX  
XX 11-AUG-2000; 2000WO-EP07865.  
XX  
XX 13-AUG-1999; 99US-0373938.  
XX

PA (NOVS ) NOVARTIS AG.  
PA (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.

PI Hallenbeck PL, Chen CT;

XX WPI; 2001-202871/20.  
DR N-PSDB; AAF60336.

XX Adenoviral vector for treating tumors and disorders associated with  
PT angiogenesis, such as cancer, arthritis, and psoriasis, comprises a DNA  
PT sequence encoding an angiogenic inhibitor, particularly endostatin -  
XX  
PS Example 1; Fig 1B; 59pp; English.

XX The nucleotide sequence encoding this protein was used in the  
CC construction of an adenoviral vector which includes a DNA sequence  
CC encoding endostatin. The adenoviral vector is useful for expressing  
CC endostatin in a mammalian cell such as an A549 or Hep3B cell. It is  
CC useful for treating other diseases and disorders associated with  
CC angiogenesis, such as neovascular diseases of the eye, including diabetic  
CC retinopathy, cardiovascular disease, arthritis, psoriasis, cerebral  
CC oedema and intravascular coagulopathy (Kasabach-Merritt syndrome). The  
CC vector inhibits, prevents or destroys the growth of tumours by  
CC preventing the formation of blood vessels in tumours, such as lymphoma  
CC and leukaemia.

XX Sequence 207 AA;

Query Match 99.7%; Score 920; DB 22; Length 207;  
Best Local Similarity 99.4%; Pred. No. 7.6e-102;  
Matches 174; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 HTHQDFQPVHLVALNTPLSGMVGIRGADFQCFQQAQAVGLSGTFRFLSSRLQDLYSI 60  
DB 24 HTHQDFQPVHLVALNTPLSGMVGIRGADFQCFQQAQAVGLSGTFRFLSSRLQDLYSI 83  
QY 61 VRRADRGSPVIVNLKDEVLSPSWDSLFSGSQGLQPGARIFSFDRDLRHPAPQKSVW 120  
DB 84 VRRADRGSPVIVNLKDEVLSPSWDSLFSGSQGLQPGARIFSFDRDLRHPAPQKSVW 143  
QY 121 HGSDPSGRRLMESYCETWRTTGTATGQASSLLSGLLEQKAASCHNSYIVLCIE 175  
DB 144 HGSDPSGRRLMESYCETWRTTGTATGQASSLLSGLLEQKAASCHNSYIVLCIE 198

RESULT 11  
AAY08691  
ID AAY08691 standard; Protein; 218 AA.

XX AAY08691;

XX 10-AUG-1999 (first entry)

XX Murine gene therapy peptide construct SP-Flag-Endo.

DE Plasmidogen; murine; angiostatin; endostatin; gene therapy; vector;  
XX anti-angiogenic; attenuation; cytostatic; anti-diabetic; ophthalmology;  
KW tumour growth; solid tumour; diabetic retinopathy; retina; construct.

XX Mus sp.  
OS Synthetic.

PN WO9926480-A1.  
XX  
PD 03-JUN-1999.  
XX  
PF 20-NOV-1998; 98WO-US24950.  
XX  
PR 20-NOV-1997; 97US-0975424.  
XX  
PA (GENE-) GENETIX PHARM INC.  
PA (MASI ) MASSACHUSETTS INST TECHNOLOGY.  
XX  
PI Bachelot T, Leboulch P, Pawliuk RJ;  
XX  
DR WPI; 1999-357696/30.  
DR N-PSDB; AAX77717.  
XX  
PT Anti-angiogenic gene therapy vectors  
XX  
PS Example 1; Page 69; 83pp; English.  
XX  
CC This invention describes a novel viral gene therapy vector comprising a  
CC nucleic acid molecule encoding an anti-angiogenic polypeptide chosen  
CC from human or murine angiostatin, human or murine endostatin and  
CC angiogenesis-inhibiting fusions and fragments, where the viral vector is  
CC sufficiently attenuated for use in human gene therapy. The products of  
CC the invention have anti-angiogenic, cytostatic, anti-diabetic and  
CC ophthalmological activity. The vector is used in gene therapy for  
CC inhibiting tumour growth in humans harbouring a solid tumour. The vector  
CC expresses an anti-angiogenic polypeptide. An additional use comprises  
CC treatment of diabetic retinopathy, where the anti-angiogenic polypeptide  
CC inhibits angiogenesis in the vicinity of the retina. The vector is  
CC administered to cells ex vivo and then administered to the patient.  
XX  
SQ Sequence 218 AA;  
  
Query Match 99.7%; Score 920; DB 20; Length 218;  
Best Local Similarity 99.4%; Pred. No. 8.2e-102;  
Matches 174; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 HTHQDFQPVHLVALNTPLSGMGRGIRGADFQCFQQAARAVGLSGTFRFLSSRLQDLYSI 60  
Db 35 HTHQDFQPVHLVALNTPLSGMGRGIRGADFQCFQQAARAVGLSGTFRFLSSRLQDLYSI 94  
  
QY 61 VRRADRGSPVIVNLKDEVLSFSGSQQLPGARIFSFDRDLRHPAPQKSVW 120  
Db 95 VRRADRGSPVIVNLKDEVLSFSGSQQLPGARIFSFDRDLRHPAPQKSVW 154  
  
QY 121 HGSDPSGRRLMESYCETWRTTETTGATGQASSLLSGRLLEKQAASCHNSYIVLCIE 175  
Db 155 HGSDPSGRRLMESYCETWRTTETTGATGQASSLLSGRLLEKQAASCHNSYIVLCIE 209  
  
RESULT 12  
AAY08692  
ID AAY08692 standard; Protein; 580 AA.  
XX  
AC AAY08692;  
XX  
DT 10-AUG-1999 (first entry)  
XX  
DE Murine gene therapy peptide construct SP-K1-K2-K3-K4-Flag-Endo.  
XX  
KW Plasminogen; murine; angiostatin; endostatin; gene therapy; vector;  
KW anti-angiogenic; attenuation; cytostatic; anti-diabetic; ophthalmology;  
KW tumour growth; solid tumour; diabetic retinopathy; retina; construct.  
XX  
OS Mus sp.  
OS Synthetic.  
XX  
PN WO9926480-A1.  
XX  
PD 03-JUN-1999.

PF 20-NOV-1998; 98WO-US24950.  
XX  
PR 20-NOV-1997; 97US-0975424.  
XX  
PA (GENE-) GENETIX PHARM INC.  
PA (MASI ) MASSACHUSETTS INST TECHNOLOGY.  
XX  
PI Bachelot T, Leboulch P, Pawliuk RJ;  
XX  
DR WPI; 1999-357696/30.  
DR N-PSDB; AAX77718.  
XX  
PT Anti-angiogenic gene therapy vectors  
XX  
PS Example 1; Page 72-74; 83pp; English.  
XX  
CC This invention describes a novel viral gene therapy vector comprising a  
CC nucleic acid molecule encoding an anti-angiogenic polypeptide chosen  
CC from human or murine angiostatin, human or murine endostatin and  
CC angiogenesis-inhibiting fusions and fragments, where the viral vector is  
CC sufficiently attenuated for use in human gene therapy. The products of  
CC the invention have anti-angiogenic, cytostatic, anti-diabetic and  
CC ophthalmological activity. The vector is used in gene therapy for  
CC inhibiting tumour growth in humans harbouring a solid tumour. The vector  
CC expresses an anti-angiogenic polypeptide. An additional use comprises  
CC treatment of diabetic retinopathy, where the anti-angiogenic polypeptide  
CC inhibits angiogenesis in the vicinity of the retina. The vector is  
CC administered to cells ex vivo and then administered to the patient.  
XX  
SQ Sequence 580 AA;  
  
Query Match 99.7%; Score 920; DB 20; Length 580;  
Best Local Similarity 99.4%; Pred. No. 3.4e-101;  
Matches 174; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 HTHQDFQPVHLVALNTPLSGMGRGIRGADFQCFQQAARAVGLSGTFRFLSSRLQDLYSI 60  
Db 397 HTHQDFQPVHLVALNTPLSGMGRGIRGADFQCFQQAARAVGLSGTFRFLSSRLQDLYSI 456  
  
QY 61 VRRADRGSPVIVNLKDEVLSFSGSQQLPGARIFSFDRDLRHPAPQKSVW 120  
Db 457 VRRADRGSPVIVNLKDEVLSFSGSQQLPGARIFSFDRDLRHPAPQKSVW 516  
  
QY 121 HGSDPSGRRLMESYCETWRTTETTGATGQASSLLSGRLLEKQAASCHNSYIVLCIE 175  
Db 517 HGSDPSGRRLMESYCETWRTTETTGATGQASSLLSGRLLEKQAASCHNSYIVLCIE 571  
  
RESULT 13  
AAY25114  
ID AAY25114 standard; Protein; 684 AA.  
XX  
AC AAY25114;  
XX  
DT 25-AUG-1999 (first entry)  
XX  
DE Mouse alpha1 (XVIII) collagen protein.  
XX  
KW Alpha1(XVIII) collagen; mimetic; endostatin; atomic coordinate; library;  
KW anti-angiogenic; heparin binding domain; receptor binding domain; mimic;  
KW alpha-helix A domain; carbohydrate recognition domain; CRD domain;  
KW treatment; angiogenesis; tumour; murine.  
XX  
OS Mus sp.  
XX  
PN WO9931616-A1.  
XX  
PD 24-JUN-1999.  
XX  
PF 16-DEC-1998; 98WO-US26783.  
XX  
PR 16-DEC-1997; 97US-0069727.  
XX

PA (HARD ) HARVARD COLLEGE.  
XX Hohenester E, Olsen BR, Sasaki T, Timpl R;  
PI WPI; 1999-395243/33.  
XX Identifying mimetics of mammalian endostatin  
PT Disclosure; Fig 5A-C; 75pp; English.  
XX  
XX This invention describes a novel method for identifying mimetics of  
CC mammalian endostatin. The method comprises identifying a compound  
CC having atomic coordinates with non-trivial similarity to selected  
CC coordinates of atoms of a mammalian endostatin involves (a) providing  
CC a library of atomic coordinates of compounds in a library of candidate  
CC compounds, (b) comparing the library of atomic coordinates to the  
CC selected coordinates of a mammalian endostatin and (c) selecting from the  
CC library at least one candidate compound on the basis of selection  
CC criteria which include similarities between the atomic coordinates of the  
CC selected candidate compound and the atomic coordinates of the mammalian  
CC endostatin. The invention also describes the use of an anti-angiogenic  
CC fragment of endostatin comprising a domain selected from a heparin  
CC binding domain, a receptor binding domain, and exposed on alpha-helix A  
CC domain, and a carbohydrate recognition domain (CRD) domain. The methods  
CC can be used for designing and selecting endostatin mimics. The compounds  
CC identified can be used for treating undesired angiogenesis, e.g. tumours.  
CC This sequence represents mouse alpha1(xviii) collagen which is used in  
CC the description of the method.  
XX  
SQ Sequence 684 AA;  
  
Query Match 99.7%; Score 920; DB 20; Length 684;  
Best Local Similarity 99.4%; Pred. No. 4.4e-101;  
Matches 174; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 1 HTHQDFQPVHLHLVALNTPLSGMGRGIRGADFCFQQAARAVGLSGTFRFLSSRLQDLYSI 60  
Db |||||||  
502 HTHQDFQPVHLHLVALNTPLSGMGRGIRGADFCFQQAARAVGLSGTFRFLSSRLQDLYSI 561  
Qy 61 VRRADRGSVPIVNLKDEVLSFSGSQQLQPGARIFSPDGRDVLHRHPAWPKSVW 120  
Db |||||||  
562 VRRADRGSVPIVNLKDEVLSFSGSQQLQPGARIFSPDGRDVLHRHPAWPKSVW 621  
Qy 121 HGSDPSGRRLMESYCTWRTTGTATGQASSLLSGRLLLEQKAASCHNSYIVLCIE 175  
Db |||||||  
622 HGSDPSGRRLMESYCTWRTTGTATGQASSLLSGRLLLEQKAASCHNSYIVLCIE 676  
  
RESULT 14  
AAW26328  
ID AAW26328 standard; Protein; 1288 AA.  
XX  
AC AAW26328;  
XX  
DT 19-NOV-1997 (first entry)  
XX  
DE Mouse alpha-1 collagen (XVIII).  
XX  
KW Alpha-1 collagen; type XVIII collagen; cartilage degeneration.  
XX  
OS Mus musculus.  
XX  
FH Key Location/Qualifiers  
FT Peptide 303..308  
FT /label= GXYGX'Y' \_motif  
FT Peptide 309..314  
FT /label= GXYGX'Y' \_motif  
FT Peptide 315..320  
FT /label= GXYGX'Y' \_motif  
FT Peptide 321..326  
FT /label= GXYGX'Y' \_motif  
FT Peptide 337..342  
FT /label= GXYGX'Y' \_motif  
FT

FT Peptide 343..348  
FT /label= GXYGX'Y' \_motif  
FT Peptide 349..354  
FT /label= GXYGX'Y' \_motif  
FT Peptide 355..360  
FT /label= GXYGX'Y' \_motif  
FT Peptide 361..366  
FT /label= GXYGX'Y' \_motif  
FT Peptide 367..372  
FT /label= GXYGX'Y' \_motif  
FT Peptide 373..378  
FT /label= GXYGX'Y' \_motif  
FT Peptide 379..384  
FT /label= GXYGX'Y' \_motif  
FT Peptide 385..390  
FT /label= GXYGX'Y' \_motif  
FT Peptide 396..401  
FT /label= GXYGX'Y' \_motif  
FT Peptide 402..407  
FT /label= GXYGX'Y' \_motif  
FT Peptide 435..440  
FT /label= GXYGX'Y' \_motif  
FT Peptide 441..446  
FT /label= GXYGX'Y' \_motif  
FT Peptide 447..452  
FT /label= GXYGX'Y' \_motif  
FT Peptide 453..458  
FT /label= GXYGX'Y' \_motif  
FT Peptide 459..464  
FT /label= GXYGX'Y' \_motif  
FT Peptide 470..475  
FT /label= GXYGX'Y' \_motif  
FT Peptide 476..481  
FT /label= GXYGX'Y' \_motif  
FT Peptide 482..487  
FT /label= GXYGX'Y' \_motif  
FT Peptide 488..493  
FT /label= GXYGX'Y' \_motif  
FT Peptide 494..499  
FT /label= GXYGX'Y' \_motif  
FT Peptide 500..505  
FT /label= GXYGX'Y' \_motif  
FT Peptide 506..511  
FT /label= GXYGX'Y' \_motif  
FT Peptide 512..517  
FT /label= GXYGX'Y' \_motif  
FT Peptide 518..523  
FT /label= GXYGX'Y' \_motif  
FT Peptide 524..529  
FT /label= GXYGX'Y' \_motif  
FT Peptide 530..535  
FT /label= GXYGX'Y' \_motif  
FT Peptide 536..541  
FT /label= GXYGX'Y' \_motif  
FT Peptide 542..547  
FT /label= GXYGX'Y' \_motif  
FT Peptide 548..553  
FT /label= GXYGX'Y' \_motif  
FT Peptide 580..585  
FT /label= GXYGX'Y' \_motif  
FT Peptide 586..591  
FT /label= GXYGX'Y' \_motif  
FT Peptide 592..597  
FT /label= GXYGX'Y' \_motif  
FT Peptide 598..603  
FT /label= GXYGX'Y' \_motif  
FT Peptide 604..609  
FT /label= GXYGX'Y' \_motif  
FT Peptide 610..615  
FT /label= GXYGX'Y' \_motif  
FT Peptide 616..621  
FT /label= GXYGX'Y' \_motif  
FT Peptide 622..627



FT /label= GYGX'Y'\_motif  
FT 628..633  
FT /label= GYGX'Y'\_motif  
FT 634..639  
FT /label= GYGX'Y'\_motif  
FT 640..665  
FT /label= GYGX'Y'\_motif  
FT 657..662  
FT /label= GYGX'Y'\_motif  
FT 677..682  
FT /label= GYGX'Y'\_motif  
FT 683..688  
FT /label= GYGX'Y'\_motif  
FT 689..694  
FT /label= GYGX'Y'\_motif  
FT 695..700  
FT /label= GYGX'Y'\_motif  
FT 707..712  
FT /label= GYGX'Y'\_motif  
FT 713..718  
FT /label= GYGX'Y'\_motif  
FT 735..740  
FT /label= GYGX'Y'\_motif  
FT 741..746  
FT /label= GYGX'Y'\_motif  
FT 747..752  
FT /label= GYGX'Y'\_motif  
FT 759..764  
FT /label= GYGX'Y'\_motif  
FT 765..770  
FT /label= GYGX'Y'\_motif  
FT 771..776  
FT /label= GYGX'Y'\_motif  
FT 787..792  
FT /label= GYGX'Y'\_motif  
FT 793..798  
FT /label= GYGX'Y'\_motif  
FT 799..804  
FT /label= GYGX'Y'\_motif  
FT 815..820  
FT /label= GYGX'Y'\_motif  
FT 821..826  
FT /label= GYGX'Y'\_motif  
FT 827..832  
FT /label= GYGX'Y'\_motif  
FT 833..838  
FT /label= GYGX'Y'\_motif  
FT 839..844  
FT /label= GYGX'Y'\_motif  
FT 845..850  
FT /label= GYGX'Y'\_motif  
FT 863..868  
FT /label= GYGX'Y'\_motif  
FT 869..874  
FT /label= GYGX'Y'\_motif  
FT 875..880  
FT /label= GYGX'Y'\_motif  
FT 891..896  
FT /label= GYGX'Y'\_motif  
FT 897..902  
FT /label= GYGX'Y'\_motif  
FT 903..908  
FT /label= GYGX'Y'\_motif  
FT 911..916  
FT /label= GYGX'Y'\_motif  
FT 917..922  
FT /label= GYGX'Y'\_motif  
FT 928..933  
FT /label= GYGX'Y'\_motif  
FT 934..939  
FT /label= GYGX'Y'\_motif  
FT 956..961  
FT /label= GYGX'Y'\_motif

Peptide 962..967  
/label= GYGX'Y'\_motif  
Peptide 968..973  
/label= GYGX'Y'\_motif  
Peptide 1126..1131  
/label= GYGX'Y'\_motif  
Peptide 1145..1150  
/label= GYGX'Y'\_motif  
Peptide 1193..1198  
/label= GYGX'Y'\_motif  
  
XX US5643783-A.  
PN  
XX  
PD 01-JUL-1997.  
XX  
PF 01-DEC-1993; 93US-0159784.  
XX  
PR 01-DEC-1993; 93US-0159784.  
XX  
PA (HARD ) HARVARD COLLEGE.  
XX  
PI Oh SP, Olsen BR;  
XX  
DR WPI; 1997-350247/32.  
DR N-PSDB; AAT84485.  
XX  
PT Nucleic acid encoding human alpha-1 collagen - for production of  
PT recombinant alpha-1 collagen, for use in the treatment of cartilage  
PT degeneration  
XX  
PS Disclosure; Fig 2; 35pp; English.  
  
Query Match 99.7%; Score 920; DB 18; Length 1288;  
Best Local Similarity 99.4%; Pred. No. 1.1e-100;  
Matches 174; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 HTHQDFQPVHLVALNTPLSGMGRGADFCFQQARAVGLSGTFRFLSSRLQDLYSI 60  
Db 1105 HTHQDFQPVHLVALNTPLSGMGRGADFCFQQARAVGLSGTFRFLSSRLQDLYSI 1164  
  
QY 61 VRRADRGSPVIVNLKDEVLSPSWDSLFSGSQQLQPGARIFPSFDGRDVLRHPAPQKSW 120  
Db 1165 VRRADRGSPVIVNLKDEVLSPSWDSLFSGSQGVQVPGARIFPSFDGRDVLRHPAPQKSW 1224  
  
QY 121 HGSDPSGRRLMESYCETWRTTGTGATGQASSLLSGRLLEQKAASCHNSYIVLCIE 175  
Db 1225 HGSDPSGRRLMESYCETWRTTGTGATGQASSLLSGRLLEQKAASCHNSYIVLCIE 1279  
  
RESULT 15  
AAW92297  
ID AAW92297 standard; peptide; 1288 AA.  
XX  
AC AAW92297;  
XX  
DT 28-APR-1999 (first entry)  
XX  
DE Mouse alpha-1 (XVIII) collagen chain common sequence MO18(common)28.  
XX  
KW Human; type XVIII collagen; liver disease; cirrhosis; detection;  
KW hepatocellular carcinoma; diagnosis.  
XX  
OS Mus sp.  
XX  
PN WO9856399-A1.  
XX  
PD 17-DEC-1998.  
XX  
PF 12-JUN-1998; 98WO-US12327.  
XX  
PR 12-JUN-1997; 97US-0049369.  
XX  
PA (FIFI-) ACAD FINLAND.

PA (FIBR-) FIBROGEN INC.  
PA (INRM ) INST NAT SANTE & RECH MEDICALE.  
XX  
PI Clement B, Pihlajaniemi T, Rehn M;  
XX  
DR WPI; 1999-070292/06.  
XX  
PT Diagnosis and monitoring of liver disease by measuring collagen type  
PT XVIII levels - with elevated levels indicative of disease,  
PT especially cirrhosis or hepatocellular carcinoma  
XX  
PS Example 6; Fig 8; 56pp; English.  
XX  
CC A method has been developed for the detecting liver disease. The method  
CC comprises: (a) reacting a patient sample with antibodies (Ab) specific  
CC for collagen type XVIII (Coll18); (b) measuring the amount of Ab-antigen  
CC complex (C) formed as indicator of the amount of Coll18 present; (c)  
CC similar analysis of a non-diseased control; and (d) comparing the  
CC amounts of Coll18 in the two samples to detect presence or progression of  
CC disease. Elevated levels of Coll18 are: (i) indicative of disease,  
CC specifically cirrhosis; and (ii) predictive of the prognosis of disease,  
CC Coll18 mRNA levels and tumour size and necrosis, and survival times are  
CC significantly higher in patients with higher Coll18 levels). The method  
CC provides non-invasive, early and accurate diagnosis of liver disease.  
CC The present sequence represents the sequence common to mouse alpha-1  
CC (XVIII) collagen chain from the present invention.  
XX  
SQ Sequence 1288 AA;  
  
Query Match 99.7%; Score 920; DB 20; Length 1288;  
Best Local Similarity 99.4%; Pred. No. 1.1e-100;  
Matches 174; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 HTHQDFQPVHLVALNTPLSGGMRGIRGADFQCFQQAQAVGLSGTFRFLSRLQDLYSI 60  
Db |||||  
1106 HTHQDFQPVHLVALNTPLSGGMRGIRGADFQCFQQAQAVGLSGTFRFLSRLQDLYSI 1165  
  
QY 61 VRRADRGSVPIVNLKDEVLSFSGWDSLFSGSQQLQPGARIFSFDRDVLRHHPWPQKSVW 120  
Db |||||  
1166 VRRADRGSVPIVNLKDEVLSFSGWDSLFSGSQGVQVQPGARIFSFDRDVLRHHPWPQKSVW 1225  
  
QY 121 HGSDPSGRRRLMESYCETWRTTGTGATGQASSLLSGRLLLEQKAASCHNSYIVLCIE 175  
Db |||||  
1226 HGSDPSGRRRLMESYCETWRTTGTGATGQASSLLSGRLLLEQKAASCHNSYIVLCIE 1280  
  
Search completed: February 17, 2004, 09:56:33  
Job time : 42 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 17, 2004, 09:55:48 ; Search time 21 Seconds  
(without alignments)  
352.590 Million cell updates/sec

Title: US-09-589-777C-2\_COPY\_1\_175

Perfect score: 923  
Sequence: 1 HTHQDFQPVHLVALNTPLS.....RLLEQKAASCHNSYIVLCIE 175

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA: \*  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep: \*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep: \*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep: \*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep: \*  
5: /cgn2\_6/ptodata/1/iaa/PTUS\_COMB.pep: \*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	923	100.0	191	4	US-09-561-500-13
2	923	100.0	191	4	US-09-561-108-13
3	923	100.0	191	4	US-09-561-526-13
4	923	100.0	191	4	US-09-561-499-13
5	920	99.7	195	1	US-08-159-784-2
6	874	94.7	185	3	US-08-985-526-36
7	813	88.1	182	4	US-09-561-500-14
8	813	88.1	182	4	US-09-561-108-14
9	813	88.1	182	4	US-09-315-689-3
10	813	88.1	182	4	US-09-561-526-14
11	813	88.1	182	4	US-09-561-499-14
12	813	88.1	183	3	US-09-206-059-2
13	795	86.1	178	4	US-09-315-689-5
14	522	56.6	191	1	US-08-159-784-3
15	150	16.3	35	3	US-09-046-985-2
16	150	16.3	35	3	US-09-474-743-2
17	108	11.7	20	2	US-08-740-168A-1
18	108	11.7	20	3	US-09-349-429-1
19	108	11.7	20	4	US-09-315-689-1
20	108	11.7	20	4	US-09-174-282-1
21	101	10.9	16	3	US-09-385-442-32
22	101	10.9	22	3	US-09-046-985-7
23	101	10.9	22	3	US-09-474-743-7
24	77	8.3	439	4	US-09-252-991A-19623
25	76.5	8.3	1050	4	US-09-428-711A-16
26	76	8.2	506	4	US-09-252-991A-18165
27	75	8.1	190	3	US-09-046-985-15

28	75	8.1	190	3	US-09-474-743-15	Sequence 15, Appl
29	74.5	8.1	577	2	US-08-756-317-13	Sequence 13, Appl
30	74	8.0	6396	4	US-09-410-551B-72	Sequence 72, Appl
31	73.5	8.0	443	4	US-09-252-991A-26460	Sequence 26460, A
32	72.5	7.9	495	4	US-08-679-493A-165	Sequence 165, App
33	72	7.8	347	4	US-09-252-991A-25606	Sequence 25606, A
34	71.5	7.7	1214	2	US-08-231-193A-54	Sequence 54, Appl
35	71.5	7.7	1214	2	US-08-486-273A-54	Sequence 54, Appl
36	71.5	7.7	1214	3	US-08-480-474-54	Sequence 54, Appl
37	71.5	7.7	1214	3	US-08-940-086A-54	Sequence 54, Appl
38	71.5	7.7	1214	4	US-08-940-035A-54	Sequence 54, Appl
39	71.5	7.7	1214	4	US-08-935-105A-54	Sequence 54, Appl
40	71.5	7.7	1214	4	US-09-648-797-54	Sequence 54, Appl
41	71.5	7.7	1214	4	US-09-386-123-54	Sequence 54, Appl
42	71.5	7.7	1219	2	US-08-231-193A-50	Sequence 50, Appl
43	71.5	7.7	1219	2	US-08-486-273A-50	Sequence 50, Appl
44	71.5	7.7	1219	3	US-08-480-474-50	Sequence 50, Appl
45	71.5	7.7	1219	3	US-08-940-086A-50	Sequence 50, Appl

ALIGNMENTS

RESULT 1  
US-09-561-500-13  
; Sequence 13, Application US/09561500  
; Patent No. 6342219  
; GENERAL INFORMATION:  
; APPLICANT: Philip E. Thorpe  
; APPLICANT: Rolf A. Brekken  
; TITLE OF INVENTION: ANTIBODY COMPOSITIONS FOR SELECTIVELY INHIBITING VEGF  
; FILE REFERENCE: 4001.002500  
; CURRENT APPLICATION NUMBER: US/09/561,500  
; CURRENT FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: 60/131,432  
; PRIOR FILING DATE: 1999-04-28  
; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 13  
; LENGTH: 191  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC  
US-09-561-500-13

Query Match	100.0%;	Score	923;	DB	4;	Length	191;
Best Local Similarity	100.0%;	Pred. No.	1.9e-106;				
Matches	175;	Conservative	0;	Mismatches	0;	Indels	0; Gaps 0;
Qy	1	HTHQDFQPVHLVALNTPLSGMGRGADFCFQQAQAVGLSGTFRFLSSRLQDLYSI	60				
Db	8	HTHQDFQPVHLVALNTPLSGMGRGADFCFQQAQAVGLSGTFRFLSSRLQDLYSI	67				
Qy	61	VRRADRGSVPIVNLKDEVLSPLSPNDLFSGSGQLQPGARIFSFDRDVLRHPAWPKSVW	120				
Db	68	VRRADRGSVPIVNLKDEVLSPLSPNDLFSGSGQLQPGARIFSFDRDVLRHPAWPKSVW	127				
Qy	121	HGSDPSGRRLMESYCETWRTTGTGATGQASSLLSGRLLEQKAASCHNSYIVLCIE	175				
Db	128	HGSDPSGRRLMESYCETWRTTGTGATGQASSLLSGRLLEQKAASCHNSYIVLCIE	182				

RESULT 2  
US-09-561-108-13  
; Sequence 13, Application US/09561108  
; Patent No. 6342221  
; GENERAL INFORMATION:  
; APPLICANT: Philip E. Thorpe  
; APPLICANT: Rolf A. Brekken  
; TITLE OF INVENTION: ANTIBODY CONJUGATE COMPOSITIONS FOR SELECTIVELY INHIBITING VEGF  
; FILE REFERENCE: 4001.002584  
; CURRENT APPLICATION NUMBER: US/09/561,108

```
; CURRENT FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/131,432
; PRIOR FILING DATE: 1999-04-28
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 191
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
US-09-561-108-13

Query Match          100.0%; Score 923; DB 4; Length 191;
Best Local Similarity 100.0%; Pred. No. 1.9e-106;
Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HTHQDFQPVHLVALNTPLSGMGRGIRGADFCFQQARAVGLSGTFRAFLSSRLQDLYSI 60
Db 8 HTHQDFQPVHLVALNTPLSGMGRGIRGADFCFQQARAVGLSGTFRAFLSSRLQDLYSI 67

QY 61 VRRADRGSPVIVNLKDEVLSFSGQQLPGARIFSFDRDVLHRPAWPQKSVW 120
Db 68 VRRADRGSPVIVNLKDEVLSFSGQQLPGARIFSFDRDVLHRPAWPQKSVW 127

QY 121 HGSDPSGRRRLMESYCETWRTTETTGATGQASSLLSGRLLEQKAASCHNSYIVLCIE 175
Db 128 HGSDPSGRRRLMESYCETWRTTETTGATGQASSLLSGRLLEQKAASCHNSYIVLCIE 182

RESULT 3
US-09-561-526-13
; Sequence 13, Application US/09561526
; Patent No. 6416758
; GENERAL INFORMATION:
; APPLICANT: Philip E. Thorpe
; APPLICANT: Rolf A. Brekken
; TITLE OF INVENTION: ANTIBODY CONJUGATE KITS FOR SELECTIVELY INHIBITING VEGF
; FILE REFERENCE: 4001.002586
; CURRENT APPLICATION NUMBER: US/09/561,526
; CURRENT FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/131,432
; PRIOR FILING DATE: 1999-04-28
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 191
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
US-09-561-526-13

Query Match          100.0%; Score 923; DB 4; Length 191;
Best Local Similarity 100.0%; Pred. No. 1.9e-106;
Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HTHQDFQPVHLVALNTPLSGMGRGIRGADFCFQQARAVGLSGTFRAFLSSRLQDLYSI 60
Db 8 HTHQDFQPVHLVALNTPLSGMGRGIRGADFCFQQARAVGLSGTFRAFLSSRLQDLYSI 67

QY 61 VRRADRGSPVIVNLKDEVLSFSGQQLPGARIFSFDRDVLHRPAWPQKSVW 120
Db 68 VRRADRGSPVIVNLKDEVLSFSGQQLPGARIFSFDRDVLHRPAWPQKSVW 127

QY 121 HGSDPSGRRRLMESYCETWRTTETTGATGQASSLLSGRLLEQKAASCHNSYIVLCIE 175
Db 128 HGSDPSGRRRLMESYCETWRTTETTGATGQASSLLSGRLLEQKAASCHNSYIVLCIE 182

RESULT 4
US-09-561-499-13
; Sequence 13, Application US/09561499
```

```
; Patent No. 6524583
; GENERAL INFORMATION:
; APPLICANT: Philip E. Thorpe
; APPLICANT: Rolf A. Brekken
; TITLE OF INVENTION: ANTIBODY METHODS FOR SELECTIVELY INHIBITING VEGF
; FILE REFERENCE: 4001.002582
; CURRENT APPLICATION NUMBER: US/09/561,499
; CURRENT FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/131,432
; PRIOR FILING DATE: 1999-04-28
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 191
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
US-09-561-499-13

Query Match          100.0%; Score 923; DB 4; Length 191;
Best Local Similarity 100.0%; Pred. No. 1.9e-106;
Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HTHQDFQPVHLVALNTPLSGMGRGIRGADFCFQQARAVGLSGTFRAFLSSRLQDLYSI 60
Db 8 HTHQDFQPVHLVALNTPLSGMGRGIRGADFCFQQARAVGLSGTFRAFLSSRLQDLYSI 67

QY 61 VRRADRGSPVIVNLKDEVLSFSGQQLPGARIFSFDRDVLHRPAWPQKSVW 120
Db 68 VRRADRGSPVIVNLKDEVLSFSGQQLPGARIFSFDRDVLHRPAWPQKSVW 127

QY 121 HGSDPSGRRRLMESYCETWRTTETTGATGQASSLLSGRLLEQKAASCHNSYIVLCIE 175
Db 128 HGSDPSGRRRLMESYCETWRTTETTGATGQASSLLSGRLLEQKAASCHNSYIVLCIE 182

RESULT 5
US-08-159-784-2
; Sequence 2, Application US/08159784
; Patent No. 5643783
; GENERAL INFORMATION:
; APPLICANT: Bjorn R. Olsen
; TITLE OF INVENTION: NOVEL COLLAGEN AND USES THEREOF
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 50Z or 55SX
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/159,784
; FILING DATE: December 1, 1993
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: John F. Freeman
; REGISTRATION NUMBER: 29,066
; REFERENCE/DOCKET NUMBER: 00246/170001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 2:
```









```

; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brook, David E.
; REGISTRATION NUMBER: 22,592
; REFERENCE/DOCKET NUMBER: CMCC-614
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781) 861-6240
; TELEFAX: (781) 861-9540
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 35 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-046-985-2

Query Match      16.3%; Score 150; DB 3; Length 35;
Best Local Similarity 91.2%; Pred. No. 1.3e-11;
Matches 31; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      36 QARAVGLSGTFRFLSSRLQDLYSIVRRADRGSV 69
Db      1 QARAVGLACTFRFLSSRLQDLYSIVRRADRAAV 34

```

Search completed: February 17, 2004, 09:58:59  
Job time : 22 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model  
Run on: February 17, 2004, 09:56:38 ; Search time 34 Seconds  
(without alignments)  
1077.703 Million cell updates/sec

Title: US-09-589-777C-2\_COPY\_1\_175  
Perfect score: 923  
Sequence: 1 HTHQDFQPVHLVALNTPLS.....RLLEQKAASCHNSYIVLCIE 175

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 801455 seqs, 209382283 residues

Total number of hits satisfying chosen parameters: 801455

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA:\*  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	923	100.0	191	10	US-09-998-831-13
2	923	100.0	191	12	US-10-373-561-13
3	923	100.0	207	14	US-10-080-797-3
4	920	99.7	184	12	US-10-292-418-18
5	920	99.7	184	15	US-10-131-241-46
6	920	99.7	207	12	US-10-422-934-71
7	874	94.7	185	14	US-10-036-869-36
8	813	88.1	180	15	US-10-131-241-56
9	813	88.1	181	15	US-10-131-241-55
10	813	88.1	182	10	US-09-998-831-14
11	813	88.1	182	12	US-10-373-561-14
12	813	88.1	182	15	US-10-131-241-54
13	813	88.1	182	15	US-10-042-347-3
14	813	88.1	183	9	US-09-873-676-2
15	813	88.1	183	12	US-10-292-418-4

16	813	88.1	183	14	US-10-080-797-1	Sequence 1, Appli
17	813	88.1	183	15	US-10-131-241-52	Sequence 52, Appl
18	813	88.1	682	12	US-10-264-049-3010	Sequence 3010, Ap
19	813	88.1	684	11	US-09-961-403-5	Sequence 5, Appli
20	813	88.1	1516	12	US-10-431-642-3	Sequence 3, Appli
21	813	88.1	1516	15	US-10-060-036-166	Sequence 166, App
22	803	87.0	184	12	US-10-292-418-35	Sequence 35, Appl
23	803	87.0	184	15	US-10-131-241-49	Sequence 49, Appl
24	802	86.9	180	15	US-10-131-241-47	Sequence 47, Appl
25	795	86.1	178	15	US-10-131-241-60	Sequence 60, Appl
26	795	86.1	178	15	US-10-042-347-5	Sequence 5, Appli
27	795	86.1	179	15	US-10-131-241-57	Sequence 57, Appl
28	793	85.9	184	12	US-09-938-391-4	Sequence 4, Appli
29	793	85.9	230	12	US-09-938-391-2	Sequence 2, Appli
30	294	31.9	63	9	US-09-822-540A-1	Sequence 1, Appli
31	169	18.3	31	9	US-09-822-540A-2	Sequence 2, Appli
32	108	11.7	20	9	US-09-873-676-115	Sequence 115, App
33	108	11.7	20	10	US-09-405-499-1	Sequence 1, Appli
34	108	11.7	20	10	US-09-174-516-1	Sequence 1, Appli
35	108	11.7	20	10	US-09-154-302-1	Sequence 1, Appli
36	108	11.7	20	12	US-10-351-284-1	Sequence 43, Appl
37	108	11.7	20	15	US-10-131-241-43	Sequence 1, Appli
38	108	11.7	20	15	US-10-232-316-1	Sequence 1, Appli
39	108	11.7	20	15	US-10-042-347-1	Sequence 1, Appli
40	101	10.9	16	10	US-09-766-412-32	Sequence 32, Appl
41	85	9.2	369	12	US-10-029-386-32926	Sequence 32926, A
42	78.5	8.5	403	12	US-10-369-493-3408	Sequence 3408, Ap
43	77	8.3	332	9	US-09-815-242-5659	Sequence 5659, Ap
44	77	8.3	346	9	US-09-815-242-12271	Sequence 12271, A
45	75	8.1	505	12	US-10-369-493-11297	Sequence 11297, A

ALIGNMENTS

RESULT 1  
US-09-998-831-13  
; Sequence 13, Application US/09998831  
; Patent No. US20020119153A1  
; GENERAL INFORMATION:  
; APPLICANT: Philip E. Thorpe  
; APPLICANT: Rolf A. Brekken  
; TITLE OF INVENTION: ANTIBODY CONJUGATE COMPOSITIONS FOR SELECTIVELY  
; TITLE OF INVENTION: INHIBITING VEGF  
; FILE REFERENCE: 4001.002584  
; CURRENT APPLICATION NUMBER: US/09/998,831  
; PRIOR FILING DATE: 2001-11-30  
; PRIOR APPLICATION NUMBER: 09/561,108  
; PRIOR FILING DATE: 2000-04-28  
; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 13  
; LENGTH: 191  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC  
US-09-998-831-13

Query Match	100.0%;	Score 923;	DB 10;	Length 191;
Best Local Similarity	100.0%;	Pred. No. 2e-96;		
Matches	175;	Conservative	0;	Mismatches 0; Indels 0; Gaps 0;
Qy	1	HTHQDFQPVHLVALNTPLSGGMRGIRGADFCQFQQAQAVGLSGTFRFLSSRLQDLYSI	60	
Db	8	HTHQDFQPVHLVALNTPLSGGMRGIRGADFCQFQQAQAVGLSGTFRFLSSRLQDLYSI	67	
Qy	61	VRRADRGSPVIVNLKDEVLSFSGSQQLQPGARIFSDGRDVLRHAPWPKSVW	120	
Db	68	VRRADRGSPVIVNLKDEVLSFSGSQQLQPGARIFSDGRDVLRHAPWPKSVW	127	
Qy	121	HGSDPSGRRLMESYCTWRTTGTGATGQASSLLSGRLLLEQKAASCHNSYIVLCIE	175	

Db 128 HGS D P S G R R L M E S Y C E T W R T E T T G A T G Q A S S L L S G R L L E Q K A A S C H N S Y I V L C I E 182

RESULT 2

US-10-373-561-13

; Sequence 13, Application US/10373561

; Publication No. US20030175276A1

; GENERAL INFORMATION:

; APPLICANT: Philip E. Thorpe

; APPLICANT: Rolf A. Brekken

; TITLE OF INVENTION: ANTIBODY METHODS FOR SELECTIVELY INHIBITING VEGF

; FILE REFERENCE: 4001.002582

; CURRENT APPLICATION NUMBER: US/10/373,561

; CURRENT FILING DATE: 2003-02-24

; PRIOR APPLICATION NUMBER: US/09/561,499

; PRIOR FILING DATE: 2000-04-28

; PRIOR APPLICATION NUMBER: 60/131,432

; PRIOR FILING DATE: 1999-04-28

; NUMBER OF SEQ ID NOS: 44

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 13

; LENGTH: 191

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC

US-10-373-561-13

Query Match 100.0%; Score 923; DB 12; Length 191;

Best Local Similarity 100.0%; Pred. No. 2e-96;

Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HTHQDFQPVHLVALNTPLSGMGRGIRGADFQCFCQARAVGLSGTFRFLSSRLQDLYSI 60

Db 8 HTHQDFQPVHLVALNTPLSGMGRGIRGADFQCFCQARAVGLSGTFRFLSSRLQDLYSI 67

QY 61 VRRADRGSVPIVNLKDEVLSFSGSWDSLSPWDSLFSGSQGQLQPGARIFSFDRDVLRHHPAWPKSVW 120

Db 68 VRRADRGSVPIVNLKDEVLSFSGSWDSLSPWDSLFSGSQGQLQPGARIFSFDRDVLRHHPAWPKSVW 127

QY 121 HGS D P S G R R L M E S Y C E T W R T E T T G A T G Q A S S L L S G R L L E Q K A A S C H N S Y I V L C I E 175

Db 128 HGS D P S G R R L M E S Y C E T W R T E T T G A T G Q A S S L L S G R L L E Q K A A S C H N S Y I V L C I E 182

RESULT 3

US-10-080-797-3

; Sequence 3, Application US/10080797

; Publication No. US20020183253A1

; GENERAL INFORMATION:

; APPLICANT: Campochiaro, Peter A.

; APPLICANT: Dixon, Katharine H.

; APPLICANT: Brazzell, Romulus K.

; TITLE OF INVENTION: METHOD FOR TREATING OCULAR

; TITLE OF INVENTION: NEOVASCULARIZATION

; FILE REFERENCE: 4-31881A

; CURRENT APPLICATION NUMBER: US/10/080,797

; CURRENT FILING DATE: 2002-02-21

; NUMBER OF SEQ ID NOS: 21

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 3

; LENGTH: 207

; TYPE: PRT

; ORGANISM: Mouse

US-10-080-797-3

Query Match 100.0%; Score 923; DB 14; Length 207;

Best Local Similarity 100.0%; Pred. No. 2.3e-96;

Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HTHQDFQPVHLVALNTPLSGMGRGIRGADFQCFCQARAVGLSGTFRFLSSRLQDLYSI 60

Db 24 HTHQDFQPVHLVALNTPLSGMGRGIRGADFQCFCQARAVGLSGTFRFLSSRLQDLYSI 83

QY 61 VRRADRGSVPIVNLKDEVLSFSGSQGQLQPGARIFSFDRDVLRHHPAWPKSVW 120

Db 84 VRRADRGSVPIVNLKDEVLSFSGSQGQLQPGARIFSFDRDVLRHHPAWPKSVW 143

QY 121 HGS D P S G R R L M E S Y C E T W R T E T T G A T G Q A S S L L S G R L L E Q K A A S C H N S Y I V L C I E 175

Db 144 HGS D P S G R R L M E S Y C E T W R T E T T G A T G Q A S S L L S G R L L E Q K A A S C H N S Y I V L C I E 198

RESULT 4

US-10-292-418-18

; Sequence 18, Application US/10292418

; Publication No. US20030139365A1

; GENERAL INFORMATION:

; APPLICANT: Lo, Kin-Ming

; APPLICANT: Li, Yue

; APPLICANT: Gillies, Stephen D

; TITLE OF INVENTION: Expression and Export of Angiogenesis Inhibitors as

; TITLE OF INVENTION: Immunofusins

; FILE REFERENCE: LEX-006C1

; CURRENT APPLICATION NUMBER: US/10/292,418

; CURRENT FILING DATE: 2002-11-12

; PRIOR APPLICATION NUMBER: 09/383,315

; PRIOR FILING DATE: 1999-08-25

; PRIOR APPLICATION NUMBER: US 60/097,883

; PRIOR FILING DATE: 1998-08-25

; NUMBER OF SEQ ID NOS: 54

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 18

; LENGTH: 184

; TYPE: PRT

; ORGANISM: Mus musculus

US-10-292-418-18

Query Match 99.7%; Score 920; DB 12; Length 184;

Best Local Similarity 99.4%; Pred. No. 4.2e-96;

Matches 174; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 HTHQDFQPVHLVALNTPLSGMGRGIRGADFQCFCQARAVGLSGTFRFLSSRLQDLYSI 60

Db 1 HTHQDFQPVHLVALNTPLSGMGRGIRGADFQCFCQARAVGLSGTFRFLSSRLQDLYSI 60

QY 61 VRRADRGSVPIVNLKDEVLSFSGSWDSLSPWDSLFSGSQGQLQPGARIFSFDRDVLRHHPAWPKSVW 120

Db 61 VRRADRGSVPIVNLKDEVLSFSGSWDSLSPWDSLFSGSQGQLQPGARIFSFDRDVLRHHPAWPKSVW 120

QY 121 HGS D P S G R R L M E S Y C E T W R T E T T G A T G Q A S S L L S G R L L E Q K A A S C H N S Y I V L C I E 175

Db 121 HGS D P S G R R L M E S Y C E T W R T E T T G A T G Q A S S L L S G R L L E Q K A A S C H N S Y I V L C I E 175

RESULT 5

US-10-131-241-46

; Sequence 46, Application US/10131241

; Publication No. US20030012792A1

; GENERAL INFORMATION:

; APPLICANT: Holaday, John W.

; APPLICANT: Fortier, Anne H.

; TITLE OF INVENTION: Compositions and Methods for Inhibiting Endothelial Cell Prolifer

; TITLE OF INVENTION: and Regulating Angiogenesis Using Cancer Markers

; FILE REFERENCE: 05213-0344 43170-271565

; CURRENT APPLICATION NUMBER: US/10/131,241

; CURRENT FILING DATE: 2002-07-22

; PRIOR APPLICATION NUMBER: US 09/413,049

; PRIOR FILING DATE: 1999-10-06

; PRIOR APPLICATION NUMBER: US 09/316,802

; PRIOR FILING DATE: 1999-05-21

; PRIOR APPLICATION NUMBER: US 60/086,586

; PRIOR FILING DATE: 1998-05-22

; NUMBER OF SEQ ID NOS: 65

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 46



```
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-131-241-56

Query Match      88.1%; Score 813; DB 15; Length 180;
Best Local Similarity 85.7%; Pred. No. 5.7e-84;
Matches 150; Conservative 14; Mismatches 11; Indels 0; Gaps 0;

Qy 1 HTHQDFQPVHLVALNTPLSGGMRGIRGADFQCFQQAARAVGLSGTFRFLSSRLQDLYSI 60
Db 1 HSHRDFQPVHLVALNSPLSGGMRGIRGADFQCFQQAARAVGLAGTFRFLSSRLQDLYSI 60

Qy 61 VRRADRGSVPIVNLKDEVLSFSGSQQLQPGARIFSGDGRDVLRRHPAPQKSVW 120
Db 61 VRRADRAAVPIVNLKDELLFPSSWEALFSGSEGPKPGARIFSGDKDVLRRHPTWPQKSVW 120

Qy 121 HGSDPSGRRRLMESYCETWRTTGTGATGQASSLLSGRLLLEQKAASCHNSYIVLCIE 175
Db 121 HGSDPNGRRLTESYCETWRTTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIE 175

RESULT 9
US-10-131-241-55
; Sequence 55, Application US/10131241
; Publication No. US20030012792A1
; GENERAL INFORMATION:
; APPLICANT: Holaday, John W.
; APPLICANT: Fortier, Anne H.
; TITLE OF INVENTION: Compositions and Methods for Inhibiting Endothelial Cell Prolifer
; TITLE OF INVENTION: and Regulating Angiogenesis Using Cancer Markers
; FILE REFERENCE: 05213-0344 43170-271565
; CURRENT APPLICATION NUMBER: US/10/131,241
; CURRENT FILING DATE: 2002-07-22
; PRIOR APPLICATION NUMBER: US 09/413,049
; PRIOR FILING DATE: 1999-10-06
; PRIOR APPLICATION NUMBER: US 09/316,802
; PRIOR FILING DATE: 1999-05-21
; PRIOR APPLICATION NUMBER: US 60/086,586
; PRIOR FILING DATE: 1998-05-22
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 55
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-131-241-55

Query Match      88.1%; Score 813; DB 15; Length 181;
Best Local Similarity 85.7%; Pred. No. 5.7e-84;
Matches 150; Conservative 14; Mismatches 11; Indels 0; Gaps 0;

Qy 1 HTHQDFQPVHLVALNTPLSGGMRGIRGADFQCFQQAARAVGLSGTFRFLSSRLQDLYSI 60
Db 1 HSHRDFQPVHLVALNSPLSGGMRGIRGADFQCFQQAARAVGLAGTFRFLSSRLQDLYSI 60

Qy 61 VRRADRGSVPIVNLKDEVLSFSGSQQLQPGARIFSGDGRDVLRRHPAPQKSVW 120
Db 61 VRRADRAAVPIVNLKDELLFPSSWEALFSGSEGPKPGARIFSGDKDVLRRHPTWPQKSVW 120

Qy 121 HGSDPSGRRRLMESYCETWRTTGTGATGQASSLLSGRLLLEQKAASCHNSYIVLCIE 175
Db 121 HGSDPNGRRLTESYCETWRTTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIE 175

RESULT 10
US-09-998-831-14
; Sequence 14, Application US/09998831
; Patent No. US20020119153A1
; GENERAL INFORMATION:
; APPLICANT: Philip E. Thorpe
; APPLICANT: Rolf A. Brekken
; TITLE OF INVENTION: ANTIBODY CONJUGATE COMPOSITIONS FOR SELECTIVELY
; TITLE OF INVENTION: INHIBITING VEGF
```

```
; FILE REFERENCE: 4001.002584
; CURRENT APPLICATION NUMBER: US/09/998,831
; CURRENT FILING DATE: 2001-11-30
; PRIOR APPLICATION NUMBER: 09/561,108
; PRIOR FILING DATE: 2000-04-28
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14
; LENGTH: 182
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
; OTHER INFORMATION: PEPTIDE
US-09-998-831-14

Query Match      88.1%; Score 813; DB 10; Length 182;
Best Local Similarity 85.7%; Pred. No. 5.7e-84;
Matches 150; Conservative 14; Mismatches 11; Indels 0; Gaps 0;

Qy 1 HTHQDFQPVHLVALNTPLSGGMRGIRGADFQCFQQAARAVGLSGTFRFLSSRLQDLYSI 60
Db 1 HSHRDFQPVHLVALNSPLSGGMRGIRGADFQCFQQAARAVGLAGTFRFLSSRLQDLYSI 60

Qy 61 VRRADRGSVPIVNLKDEVLSFSGSQQLQPGARIFSGDGRDVLRRHPAPQKSVW 120
Db 61 VRRADRAAVPIVNLKDELLFPSSWEALFSGSEGPKPGARIFSGDKDVLRRHPTWPQKSVW 120

Qy 121 HGSDPSGRRRLMESYCETWRTTGTGATGQASSLLSGRLLLEQKAASCHNSYIVLCIE 175
Db 121 HGSDPNGRRLTESYCETWRTTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIE 175

RESULT 11
US-10-373-561-14
; Sequence 14, Application US/10373561
; Publication No. US20030175276A1
; GENERAL INFORMATION:
; APPLICANT: Philip E. Thorpe
; APPLICANT: Rolf A. Brekken
; TITLE OF INVENTION: ANTIBODY METHODS FOR SELECTIVELY INHIBITING VEGF
; FILE REFERENCE: 4001.002582
; CURRENT APPLICATION NUMBER: US/10/373,561
; CURRENT FILING DATE: 2003-02-24
; PRIOR APPLICATION NUMBER: US/09/561,499
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/131,432
; PRIOR FILING DATE: 1999-04-28
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14
; LENGTH: 182
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
; OTHER INFORMATION: PEPTIDE
US-10-373-561-14

Query Match      88.1%; Score 813; DB 12; Length 182;
Best Local Similarity 85.7%; Pred. No. 5.7e-84;
Matches 150; Conservative 14; Mismatches 11; Indels 0; Gaps 0;

Qy 1 HTHQDFQPVHLVALNTPLSGGMRGIRGADFQCFQQAARAVGLSGTFRFLSSRLQDLYSI 60
Db 1 HSHRDFQPVHLVALNSPLSGGMRGIRGADFQCFQQAARAVGLAGTFRFLSSRLQDLYSI 60

Qy 61 VRRADRGSVPIVNLKDEVLSFSGSQQLQPGARIFSGDGRDVLRRHPAPQKSVW 120
Db 61 VRRADRAAVPIVNLKDELLFPSSWEALFSGSEGPKPGARIFSGDKDVLRRHPTWPQKSVW 120

Qy 121 HGSDPSGRRRLMESYCETWRTTGTGATGQASSLLSGRLLLEQKAASCHNSYIVLCIE 175
```



Db 121 HGSDPNGRRLTESYCETWRTPEPSATGQASSLLGGRLLGQSAASCHHAYIVLCIE 175

RESULT 12

US-10-131-241-54

; Sequence 54, Application US/10131241

; Publication No. US20030012792A1

; GENERAL INFORMATION:

; APPLICANT: Holaday, John W.

; APPLICANT: Fortier, Anne H.

; TITLE OF INVENTION: Compositions and Methods for Inhibiting Endothelial Cell Prolifer

; TITLE OF INVENTION: and Regulating Angiogenesis Using Cancer Markers

; FILE REFERENCE: 05213-0344 43170-271565

; CURRENT APPLICATION NUMBER: US/10/131,241

; CURRENT FILING DATE: 2002-07-22

; PRIOR APPLICATION NUMBER: US 09/413,049

; PRIOR FILING DATE: 1999-10-06

; PRIOR APPLICATION NUMBER: US 09/316,802

; PRIOR FILING DATE: 1999-05-21

; PRIOR APPLICATION NUMBER: US 60/086,586

; PRIOR FILING DATE: 1998-05-22

; NUMBER OF SEQ ID NOS: 65

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 54

; LENGTH: 182

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-131-241-54

Query Match 88.1%; Score 813; DB 15; Length 182;

Best Local Similarity 85.7%; Pred. No. 5.7e-84;

Matches 150; Conservative 14; Mismatches 11; Indels 0; Gaps 0;

QY 1 HTHQDFQPVHLHVALNTPLSGMGRGIRGADFCFQQAARAVGLSGTFRFLSSRLQDLYSI 60

Db 1 HSHRDFQPVHLHVALNSPLSGMGRGIRGADFCFQQAARAVGLAGTFRFLSSRLQDLYSI 60

QY 61 VRRADRGSPVIVNLKDEVLSFSGSQQLQPGARIFSFQDGRDVLRHHPWPQKSVW 120

Db 61 VRRADRAAVPIVNLKDELLFPSEALFSGEGPLKPGARIFSFQDKDVLRHPTWPQKSVW 120

QY 121 HGSDPSGRRLMESYCETWRTTETGATGQASSLLSGRLLSQKAAASCHNSYIVLCIE 175

Db 121 HGSDPNGRRLTESYCETWRTPEPSATGQASSLLGGRLLGQSAASCHHAYIVLCIE 175

RESULT 13

US-10-042-347-3

; Sequence 3, Application US/10042347

; Publication No. US20030114370A1

; GENERAL INFORMATION:

; APPLICANT: O'Reilly, Michael S.

; APPLICANT: Folkman, M. Judah

; TITLE OF INVENTION: Nucleic Acid Molecules Encoding Endostatin Protein and Peptide Fr

; TITLE OF INVENTION: Thereof

; FILE REFERENCE: 05213-0880 (43170-249874)

; CURRENT APPLICATION NUMBER: US/10/042,347

; CURRENT FILING DATE: 2002-01-11

; PRIOR APPLICATION NUMBER: US 09/315,689

; PRIOR FILING DATE: 1999-05-20

; PRIOR APPLICATION NUMBER: US 60/106,343

; PRIOR FILING DATE: 1998-10-30

; PRIOR APPLICATION NUMBER: US 09/154,302

; PRIOR FILING DATE: 1998-09-16

; PRIOR APPLICATION NUMBER: US 08/740,168

; PRIOR FILING DATE: 1996-10-22

; PRIOR APPLICATION NUMBER: US 60/005,835

; PRIOR FILING DATE: 1995-10-23

; PRIOR APPLICATION NUMBER: US 60/023,070

; PRIOR FILING DATE: 1996-08-02

; PRIOR APPLICATION NUMBER: US 60/026,263

; PRIOR FILING DATE: 1996-09-17

; NUMBER OF SEQ ID NOS: 6

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 3

; LENGTH: 182

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-042-347-3

Query Match 88.1%; Score 813; DB 15; Length 182;

Best Local Similarity 85.7%; Pred. No. 5.7e-84;

Matches 150; Conservative 14; Mismatches 11; Indels 0; Gaps 0;

QY 1 HTHQDFQPVHLHVALNTPLSGMGRGIRGADFCFQQAARAVGLSGTFRFLSSRLQDLYSI 60

Db 1 HSHRDFQPVHLHVALNSPLSGMGRGIRGADFCFQQAARAVGLAGTFRFLSSRLQDLYSI 60

QY 61 VRRADRGSPVIVNLKDEVLSFSGSQQLQPGARIFSFQDGRDVLRHHPWPQKSVW 120

Db 61 VRRADRAAVPIVNLKDELLFPSEALFSGEGPLKPGARIFSFQDKDVLRHPTWPQKSVW 120

QY 121 HGSDPSGRRLMESYCETWRTTETGATGQASSLLSGRLLSQKAAASCHNSYIVLCIE 175

Db 121 HGSDPNGRRLTESYCETWRTPEPSATGQASSLLGGRLLGQSAASCHHAYIVLCIE 175

RESULT 14

US-09-873-676-2

; Sequence 2, Application US/09873676

; Patent No. US20020077289A1

; GENERAL INFORMATION:

; APPLICANT: MacDonald, Nicholas J.

; APPLICANT: Sim, Kim L.

; TITLE OF INVENTION: Angiostatin and Endostatin Binding Proteins and Methods of Use

; FILE REFERENCE: 05213-0378 (43170-259333)

; CURRENT APPLICATION NUMBER: US/09/873,676

; CURRENT FILING DATE: 2001-06-04

; PRIOR APPLICATION NUMBER: US 60/209,065

; PRIOR FILING DATE: 2000-06-02

; PRIOR APPLICATION NUMBER: US 60/289,387

; PRIOR FILING DATE: 2001-05-08

; NUMBER OF SEQ ID NOS: 123

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 2

; LENGTH: 183

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-873-676-2

Query Match 88.1%; Score 813; DB 9; Length 183;

Best Local Similarity 85.7%; Pred. No. 5.8e-84;

Matches 150; Conservative 14; Mismatches 11; Indels 0; Gaps 0;

QY 1 HTHQDFQPVHLHVALNTPLSGMGRGIRGADFCFQQAARAVGLSGTFRFLSSRLQDLYSI 60

Db 1 HSHRDFQPVHLHVALNSPLSGMGRGIRGADFCFQQAARAVGLAGTFRFLSSRLQDLYSI 60

QY 61 VRRADRGSPVIVNLKDEVLSFSGSQQLQPGARIFSFQDGRDVLRHHPWPQKSVW 120

Db 61 VRRADRAAVPIVNLKDELLFPSEALFSGEGPLKPGARIFSFQDKDVLRHPTWPQKSVW 120

QY 121 HGSDPSGRRLMESYCETWRTTETGATGQASSLLSGRLLSQKAAASCHNSYIVLCIE 175

Db 121 HGSDPNGRRLTESYCETWRTPEPSATGQASSLLGGRLLGQSAASCHHAYIVLCIE 175

RESULT 15

US-10-292-418-4

; Sequence 4, Application US/10292418

; Publication No. US20030139365A1

; GENERAL INFORMATION:

; APPLICANT: Lo, Kin-Ming

; APPLICANT: Li, Yue

; APPLICANT: Gillies, Stephen D

; TITLE OF INVENTION: Expression and Export of Angiogenesis Inhibitors as



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 17, 2004, 09:52:13 ; Search time 21 Seconds  
(without alignments)  
801.406 Million cell updates/sec

Title: US-09-589-777C-2\_COPY\_1\_175  
Perfect score: 923  
Sequence: 1 HTHQDFQPVHLVALNTPLS.....RLLEQKAASCHNSYIVLCIE 175

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 76: \*  
1: pir1: \*  
2: pir2: \*  
3: pir3: \*  
4: pir4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	923	100.0	1774	2 B56101	collagen alpha 1(X
2	920	99.7	1315	2 A56101	collagen alpha 1(X
3	813	88.1	684	2 A53019	collagen alpha 1(X
4	535	58.0	1388	2 A53317	collagen alpha 1(X
5	370	40.1	650	2 T22002	hypothetical prote
6	85	9.2	995	2 H59432	RhoGAP protein hom
7	79.5	8.6	286	2 E97241	hypothetical prote
8	79	8.6	427	2 A53798	58K membrane-assoc
9	78.5	8.5	416	2 AD0921	probable lipopolys
10	78	8.5	1055	2 S53597	chlorophyll a/b-bi
11	77.5	8.4	474	2 T10011	probable penicilli
12	77.5	8.4	492	2 B86911	probable penicilli
13	77.5	8.4	6420	2 T30283	polyketide synthas
14	77	8.3	208	2 T45272	methyltransferase
15	77	8.3	351	2 B89781	conserved hypothet
16	77	8.3	534	2 JC7912	amino-acid N-acety
17	77	8.3	1575	2 S68448	synaptojanin, 170K
18	76.5	8.3	608	2 A97685	hypothetical prote
19	76.5	8.3	608	2 AC2910	hypothetical prote
20	76.5	8.3	996	2 G87687	hypothetical prote
21	76.5	8.3	1054	2 T30933	chitinase (EC 3.2.
22	76	8.2	309	1 E65112	hypothetical 34.6
23	76	8.2	309	2 E91140	hypothetical prote
24	76	8.2	309	2 E85985	hypothetical prote
25	75.5	8.2	4924	2 T50176	probable peptide s
26	74.5	8.1	183	4 S59318	hypothetical prote
27	74.5	8.1	3670	2 T36249	CDA peptide synthe
28	74.5	8.1	7576	2 T17428	FK506 polyketide s
29	74	8.0	244	2 C45729	sulfolipid biosynt

30	73.5	8.0	488	2 A87569	peptidase M20/M25/
31	73	7.9	477	2 T05202	pectinesterase hom
32	72.5	7.9	353	2 A98131	mocA protein [impo
33	72.5	7.9	353	2 AH3156	oxidoreductase moc
34	72.5	7.9	419	2 D42725	nitrite hydratase
35	72.5	7.9	463	2 B69994	xaa-His dipeptidas
36	72.5	7.9	739	1 VHIWEB	nucleosid prote
37	72.5	7.9	800	2 T25140	hypothetical prote
38	72	7.8	309	2 AH0906	conserved hypothet
39	72	7.8	396	2 H95935	probable aminotran
40	72	7.8	442	2 I47074	gene CD5 protein -
41	72	7.8	732	2 C84487	hypothetical prote
42	72	7.8	1955	1 AGCH	agrin precursor -
43	72	7.8	2899	2 T21546	hypothetical prote
44	71.5	7.7	204	2 B45022	CRK-I - human
45	71.5	7.7	732	2 AF0439	probable Rhs acces

ALIGNMENTS

RESULT 1

B56101  
collagen alpha 1(XVIII) chain precursor, long splice form - mouse  
N;Contains: collagen alpha 1(XVIII) chain precursor, medium splice form; endostatin  
C;Species: Mus musculus (house mouse)  
C;Date: 03-Oct-1995 #sequence revision 08-May-1998 #text\_change 31-Mar-2000  
C;Accession: B56101; C56101; S72450; S65595; PN0675; A54072; A58816  
R;Rehn, M.; Pihlajaniemi, T.  
J. Biol. Chem. 270, 4705-4711, 1995  
A;Title: Identification of three N-terminal ends of type XVIII collagen chains and tissu  
tif homologous to rat and Drosophila frizzled proteins.  
A;Reference number: A56101; MUID:95181468; PMID:7876242  
A;Accession: B56101  
A;Molecule type: mRNA  
A;Residues: 1-562 <REH1>  
A;Cross-references: GB:U11637; NID:g618429; PIDN:AAC52179.1; PID:g618430  
A;Experimental source: splice form clone PE17.24  
A;Accession: C56101  
A;Molecule type: mRNA  
A;Residues: 1-239,487-562 <REH2>  
A;Cross-references: GB:U11637; NID:g618429  
A;Experimental source: splice form clones PE8.1, PE19, PE15.2  
R;Oh, S.P.; Kamagata, Y.; Muragaki, Y.; Timmons, S.; Ooshima, A.; Olsen, B.R.  
submitted to the EMBL Data Library, August 1993  
A;Reference number: S72450  
A;Accession: S72450  
A;Molecule type: mRNA  
A;Residues: 487-1146,'L',1148-1193,'F',1195-1210,'R',1212-1512,'L',1514-1522,'F',1524-16  
A;Cross-references: EMBL:L22545; NID:G348968; PIDN:AAA19787.1; PID:G511298  
R;Oh, S.P.; Kamagata, Y.; Muragaki, Y.; Timmons, S.; Ooshima, A.; Olsen, A.B.R.  
Proc. Natl. Acad. Sci. U.S.A. 91, 4229-4233, 1994  
A;Title: Isolation and sequencing of cDNAs for proteins with multiple domains of Gly-Xaa  
A;Reference number: A58370; MUID:94240111; PMID:8183893  
A;Accession: S65595  
A;Molecule type: mRNA  
A;Residues: 487-1512,'L',1514-1522,'F',1524-1683,'V',1685-1774 <OH2>  
A;Cross-references: EMBL:L22545  
R;Abe, N.; Muragaki, Y.; Yoshioka, H.; Inoue, H.; Ninomiya, Y.  
Biochem. Biophys. Res. Commun. 196, 576-582, 1993  
A;Title: Identification of a novel collagen chain represented by extensive interruptions  
A;Reference number: PN0675; MUID:94059075; PMID:8240330  
A;Accession: PN0675  
A;Molecule type: mRNA  
A;Residues: 635-1774 <ABE>  
R;Rehn, M.; Hintikka, E.; Pihlajaniemi, T.  
J. Biol. Chem. 269, 13929-13935, 1994  
A;Title: Primary structure of the alpha1 chain of mouse type XVIII collagen, partial str  
collagen chain.  
A;Reference number: A54072; MUID:94245707; PMID:8188673  
A;Accession: A54072  
A;Molecule type: DNA; mRNA  
A;Residues: 1293-1403,'R',1405-1774 <REH3>

A;Cross-references: GB:U03714; NID:G487733; PIDN:AAA20657.1; PID:G487734  
R;O'Reilly, M.S.; Boehm, T.; Shing, Y.; Fukai, N.; Vasios, G.; Lane, W.S.; Flynn, E.; Bl  
Cell 88, 277-285, 1997  
A;Title: Endostatin: an endogenous inhibitor of angiogenesis and tumor growth.  
A;Reference number: A58816; MUID:97160848; PMID:9008168  
A;Accession: A58816  
A;Molecule type: protein  
A;Residues: 1591-1610 <ORE>  
A;Experimental source: hemangioendothelium cells  
A;Note: inhibits endothelial cell proliferation  
C;Comment: Prolines and lysines at the third position of the tripeptide repeating unit  
lated and subsequently O-glycosylated.  
C;Comment: The different splice forms of collagen alpha 1(XVIII) may be involved in peri  
C;Comment: Endostatin is released from collagen alpha 1(XVIII) chain by the action of un  
ay be useful in treating solid tumors.  
C;Genetics:  
A;Gene: MGI:Coll18a1  
A;Cross-references: MGI:71175  
A;Map position: 10:41.0  
A;Introns: 1295/3; 1310/1; 1331/1; 1345/3; 1388/3; 1437/1; 1461/3; 1505/3; 1516/3; 1599/  
A;Note: the list of introns is incomplete  
C;Superfamily: unassigned collagens  
C;Keywords: alternative splicing; angiogenesis inhibitor; chondroitin sulfate proteoglyc  
F;1-1774/Product: collagen alpha 1(XVIII) chain precursor, long splice form #status pred  
F;1-239,487-1774/Product: collagen alpha 1(XVIII) chain precursor, medium splice form #s  
F;1-24/Domain: signal sequence #status predicted <SIG>  
F;361-486/Region: frizzled similarity  
F;786-812/Domain: collagenous #status predicted <CO01>  
F;823-896/Domain: collagenous #status predicted <CO02>  
F;921-1042/Domain: collagenous #status predicted <CO03>  
F;1066-1148/Domain: collagenous #status predicted <CO04>  
F;1163-1204/Domain: collagenous #status predicted <CO05>  
F;1218-1290/Domain: collagenous #status predicted <CO06>  
F;1301-1333/Domain: collagenous #status predicted <CO07>  
F;1346-1369/Domain: collagenous #status predicted <CO08>  
F;1351-1353/Region: cell attachment (R-G-D) motif  
F;1377-1428/Domain: collagenous #status predicted <CO09>  
F;1442-1459/Domain: collagenous #status predicted <CO10>  
F;1591-1774/Product: endostatin #status predicted <EST>  
F;1598-1774/Region: multiplexin collagen carboxyl-terminal similarity  
F;354,361,947/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F;699,704,1716/Binding site: carbohydrate (Ser) (covalent) #status predicted  
F;910,913,1093/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted

Query Match 100.0%; Score 923; DB 2; Length 1774;  
Best Local Similarity 100.0%; Pred. No. 6.2e-82;  
Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HTHQDFQPVHLVALNTPLSGMGRGIRGADFCFQQAQAVGLSGTFRFLSSRLQDLYSI 60  
Db 1591 HTHQDFQPVHLVALNTPLSGMGRGIRGADFCFQQAQAVGLSGTFRFLSSRLQDLYSI 1650

QY 61 VRRADRGSPVPIVNLKDEVLSFSGSQQLQPGARIFSFDRDVLHRHPWPQKSVW 120  
Db 1651 VRRADRGSPVPIVNLKDEVLSFSGSQQLQPGARIFSFDRDVLHRHPWPQKSVW 1710

QY 121 HGSDPSGRRRLMESYCETWRTTGTGATGQASSLLSGRLLEQKAASCHNSYIVLCIE 175  
Db 1711 HGSDPSGRRRLMESYCETWRTTGTGATGQASSLLSGRLLEQKAASCHNSYIVLCIE 1765

RESULT 2  
A56101  
collagen alpha 1(XVIII) chain precursor, short splice form - mouse  
N;Contains: endostatin  
C;Species: Mus musculus (house mouse)  
C;Date: 03-Oct-1995 #sequence revision 08-May-1998 #text\_change 31-Mar-2000  
C;Accession: A56101; A58371; S72450; S65595  
R;Rehn, M.; Pihlajaniemi, T.  
J. Biol. Chem. 270, 4705-4711, 1995  
A;Title: Identification of three N-terminal ends of type XVIII collagen chains and tissu  
tif homologous to rat and Drosophila frizzled proteins.  
A;Reference number: A56101; MUID:95181468; PMID:7876242

A;Accession: A56101  
A;Molecule type: mRNA  
A;Residues: 1-103 <REH1>  
A;Cross-references: GB:U11636; NID:G618427; PIDN:AAC52178.1; PID:G618428  
R;Rehn, M.; Pihlajaniemi, T.  
Proc. Natl. Acad. Sci. U.S.A. 91, 4234-4238, 1994  
A;Title: Alpha1(XVIII), a collagen chain with frequent interruptions in the collagenous  
A;Reference number: A58371; MUID:94240112; PMID:8183894  
A;Accession: A58371  
A;Molecule type: mRNA  
A;Residues: 1-928 <REH2>  
A;Cross-references: GB:L16898; NID:G404754; PIDN:AAA37434.1; PID:G553894  
R;Oh, S.P.; Kamagata, Y.; Muragaki, Y.; Timmons, S.; Ooshima, A.; Olsen, B.R.  
submitted to the EMBL Data Library, August 1993  
A;Reference number: S72450  
A;Accession: S72450  
A;Molecule type: mRNA  
A;Residues: 28-687, 'L', 689-734, 'F', 736-751, 'R', 753-1315 <OHW>  
A;Cross-references: EMBL:L22545; NID:G348968; PIDN:AA19787.1; PID:G511298  
R;Oh, S.P.; Kamagata, Y.; Muragaki, Y.; Timmons, S.; Ooshima, A.; Olsen, A.B.R.  
Proc. Natl. Acad. Sci. U.S.A. 91, 4229-4233, 1994  
A;Title: Isolation and sequencing of cDNAs for proteins with multiple domains of Gly-Xaa  
A;Reference number: A58370; MUID:94240111; PMID:8183893  
A;Accession: S65595  
A;Molecule type: mRNA  
A;Residues: 28-1315 <OHS>  
A;Cross-references: EMBL:L22545  
C;Comment: Prolines and lysines at the third position of the tripeptide repeating unit (l  
lated and subsequently O-glycosylated.  
C;Comment: The different splice forms of collagen alpha 1(XVIII) may be involved in peri  
C;Comment: Endostatin is released from collagen alpha 1(XVIII) chain by the action of un  
ay be useful in treating solid tumors.  
C;Genetics:  
A;Gene: MGI:Coll18a1  
A;Cross-references: MGI:71175  
A;Map position: 10:41.0  
C;Superfamily: unassigned collagens  
C;Keywords: alternative splicing; angiogenesis inhibitor; chondroitin sulfate proteoglyc  
F;1-25/Domain: signal sequence #status predicted <SIG>  
F;24-235/Region: thrombospondin amino-terminal similarity  
F;26-1315/Product: collagen alpha 1(XVIII) chain, short splice form #status predicted <M  
F;327-353/Domain: collagenous #status predicted <CO1>  
F;364-437/Domain: collagenous #status predicted <CO2>  
F;462-583/Domain: collagenous #status predicted <CO3>  
F;607-689/Domain: collagenous #status predicted <CO4>  
F;704-745/Domain: collagenous #status predicted <CO5>  
F;759-831/Domain: collagenous #status predicted <CO6>  
F;842-874/Domain: collagenous #status predicted <CO7>  
F;887-910/Domain: collagenous #status predicted <CO8>  
F;892-894/Region: cell attachment (R-G-D) motif  
F;918-969/Domain: collagenous #status predicted <CO9>  
F;983-1000/Domain: collagenous #status predicted <CO10>  
F;1132-1315/Product: endostatin #status predicted <EST>  
F;1139-1315/Region: multiplexin collagen carboxyl-terminal similarity  
F;126,488/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F;172-228/Disulfide bonds: #status predicted  
F;240,245,1257/Binding site: carbohydrate (Ser) (covalent) #status predicted  
F;451,454,594/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted

Query Match 99.7%; Score 920; DB 2; Length 1315;  
Best Local Similarity 99.4%; Pred. No. 8.5e-82;  
Matches 174; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 HTHQDFQPVHLVALNTPLSGMGRGIRGADFCFQQAQAVGLSGTFRFLSSRLQDLYSI 60  
Db 1132 HTHQDFQPVHLVALNTPLSGMGRGIRGADFCFQQAQAVGLSGTFRFLSSRLQDLYSI 1191

QY 61 VRRADRGSPVPIVNLKDEVLSFSGSQQLQPGARIFSFDRDVLHRHPWPQKSVW 120  
Db 1192 VRRADRGSPVPIVNLKDEVLSFSGSQQLQPGARIFSFDRDVLHRHPWPQKSVW 1251

QY 121 HGSDPSGRRRLMESYCETWRTTGTGATGQASSLLSGRLLEQKAASCHNSYIVLCIE 175

Db 1252 HGSDPSGRRLMESYCETWRTTGTGATGQASSLLSGRLLEQKAASCHNSYIVLCIE 1306

RESULT 3

A53019

collagen alpha 1(XVIII) chain - human (fragment)

N;Contains: endostatin

C;Species: Homo sapiens (man)

C;Date: 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 31-Mar-2000

C;Accession: A53019

R;Oh, S.P.; Warman, M.L.; Seldin, M.F.; Cheng, S.D.; Knoll, J.H.M.; Timmons, S.; Olsen, Genomics 19, 494-499, 1994

A;Title: Cloning of cDNA and genomic DNA encoding human type XVIII collagen and localization

A;Reference number: A53019; MUID:94245237; PMID:8188291

A;Accession: A53019

A;Molecule type: mRNA

A;Residues: 1-684 <OHA>

A;Cross-references: GB:L22548; NID:G348908; PIDN:AAA51864.1; PID:G562794

A;Note: the cited accession number, L25548, is not in Genbank release 103

A;Note: in the authors' translation, 482-Gly is not shown, residues 483-490 are shifted

C;Comment: Prolines and lysines at the third position of the tripeptide repeating unit (labeled and subsequently O-glycosylated).

C;Comment: Different splice forms of collagen alpha 1(XVIII) may be involved in perivascular

C;Comment: Endostatin is released from collagen alpha 1(XVIII) chain by the action of un-

ay be useful in treating solid tumors.

C;Genetics:

A;Gene: GDB:COL18A1

A;Cross-references: GDB:138752; OMIM:120328

A;Map position: 21q22.3-21q22.3

C;Superfamily: unassigned collagens

C;Keywords: alternative splicing; angiogenesis inhibitor; chondroitin sulfate proteoglyc

F;1-684/Product: collagen alpha 1(XVIII) chain (fragment) #status predicted <MAT>

F;1-59/Domain: collagenous (fragment) #status predicted <CO4>

F;74-115/Domain: collagenous #status predicted <CO5>

F;129-201/Domain: collagenous #status predicted <CO6>

F;212-244/Domain: collagenous #status predicted <CO7>

F;257-278/Domain: collagenous #status predicted <CO8>

F;262-264/Region: cell attachment (R-G-D) motif

F;286-340/Domain: collagenous #status predicted <CO9>

F;354-371/Domain: collagenous #status predicted <CO10>

F;502-684/Product: endostatin #status predicted <EST>

F;509-684/Region: multiplexin collagen carboxyl-terminal similarity

Query Match 88.1%; Score 813; DB 2; Length 684;

Best Local Similarity 85.7%; Pred. No. 1.2e-71;

Matches 150; Conservative 14; Mismatches 11; Indels 0; Gaps 0;

Qy 1 HTHQDFQPVHLVALNTPLSGMGRGADFCQFQQAQAVGLSGTFRFLSSRLQDLYSI 60

Db 502 HSHRDFQPVHLVALNSPLSGMGRGADFCQFQQAQAVGLAGTFRFLSSRLQDLYSI 561

Qy 61 VRRADRGSPVIVNLKDEVLSFSGSQGLQPGARIFSDGRDVLRHHPAWPKSVW 120

Db 562 VRRADRAAVPIVNLKDELLFSPWEALFSGSEGLKPGARIFSDGKDVLRHPTWPQKSVW 621

Qy 121 HGSDPSGRRLMESYCETWRTTGTGATGQASSLLSGRLLEQKAASCHNSYIVLCIE 175

Db 622 HGSDPENGRLTESYCETWRTTEAPSATGQASSLLGRLLGQSAASCHHAYIVLCIE 676

RESULT 4

A53317

collagen alpha 1(XV) chain precursor - human

N;Alternate names: procollagen alpha 1(XV) chain

C;Species: Homo sapiens (man)

C;Date: 07-Jul-1995 #sequence\_revision 07-Jul-1995 #text\_change 31-Mar-2000

C;Accession: A53317; A53146; S28778

R;Kivirikko, S.; Heinaemaeki, P.; Rehn, M.; Honkanen, N.; Myers, J.C.; Pihlajaniemi, T. J. Biol. Chem. 269, 4773-4779, 1994

A;Title: Primary structure of the alpha1 chain of human type XV collagen and exon-intron

A;Reference number: A53317; MUID:94148920; PMID:8106446

A;Accession: A53317

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-1388 <KIV>

A;Cross-references: GB:L25280

A;Note: nucleotide sequence and conceptual translation not complete

R;Muragaki, Y.; Abe, N.; Ninomiya, Y.; Olsen, B.R.; Ooshima, A. J. Biol. Chem. 269, 4042-4046, 1994

A;Title: The human alpha1(XV) collagen chain contains a large amino-terminal non-triple

A;Reference number: A53146; MUID:94140817; PMID:8307960

A;Accession: A53146

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-9,'S',11-48,'V',50-94,'A',96-149,'A',151-203,'V',205-408,'A',410-569 <MUR>

A;Cross-references: GB:D21230; NID:G415605; PIDN:BAA04762.1; PID:dl005294; PID:G460703

R;Myers, J.C.; Kivirikko, S.; Gordon, M.K.; Dion, A.S.; Pihlajaniemi, T. Proc. Natl. Acad. Sci. U.S.A. 89, 10144-10148, 1992

A;Title: Identification of a previously unknown human collagen chain, alpha1(XV), charac

A;Reference number: S28778; MUID:93066196; PMID:1279671

A;Accession: S28778

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 544-640,'P',642-811,'P',813-1252 <MYE>

C;Genetics:

A;Gene: GDB:COL15A1

A;Cross-references: GDB:132578; OMIM:120325

A;Map position: 9q21-9q22

C;Superfamily: unassigned collagens

F;1-22/Domain: signal sequence #status predicted <SIG>

F;23-1388/Product: collagen alpha 1(XV) chain #status predicted <MAT>

F;1216-1388/Region: multiplexin collagen carboxyl-terminal similarity

Query Match 58.0%; Score 535; DB 2; Length 1388;

Best Local Similarity 60.4%; Pred. No. 5.2e-44;

Matches 102; Conservative 23; Mismatches 40; Indels 4; Gaps 1;

Qy 7 QPVLHLVALNTPLSGMGRGADFCQFQQAQAVGLSGTFRFLSSRLQDLYSIVRRADR 66

Db 1215 KPALHLAALNMPFSGDIR----ADFQCFKQARAAGLSTYRAFLSSHLQDLSTIVRKAER 1270

Qy 67 GSVPIVNLKDEVLSFSGSQGLQPGARIFSDGRDVLRHHPAWPKSVWHGSDPS 126

Db 1271 YSLPIVNLKGQVLFNNWDSIFSGHGGQFNHPIVPSFDGRDITMDPSWPQKVIWHGSSPH 1330

Qy 127 GRRLMESYCETWRTTGTGATGQASSLLSGRLLEQKAASCHNSYIVLCIE 175

Db 1331 GVRLVDNYCEAWRTADTAVTGLASPLSTGKILDQKAYSCANRLIVLCIE 1379

RESULT 5

T22002

hypothetical protein F39H11.4 - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C;Accession: T22002

R;White, S.

submitted to the EMBL Data Library, October 1996

A;Reference number: Z19500

A;Accession: T22002

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-650 <WIL>

A;Cross-references: EMBL:Z81079; PIDN:CAB03084.1; GSPDB:GN00019; CESP:F39H11.4

A;Experimental source: clone F39H11

C;Genetics:

A;Gene: CESP:F39H11.4

A;Map position: 1

A;Introns: 109/3; 154/1; 357/1; 420/3; 464/3; 566/2; 594/1; 628/3

Query Match 40.1%; Score 370; DB 2; Length 650;

Best Local Similarity 44.4%; Pred. No. 3.2e-28;

Matches 75; Conservative 26; Mismatches 62; Indels 6; Gaps 4;

Qy 9 VLHLVALNTPLSGMGRGADFCQFQQAQAVGLSGTFRFLSSRLQDLYSIVRRDRGS 68

Db 1331 GVRLVDNYCEAWRTADTAVTGLASPLSTGKILDQKAYSCANRLIVLCIE 1379

Db 464 VIHMIALSQPFSGNLHGLRGADLQCYREARAAGYTTTFRAMLSSNVQDLVRIHVSVDFT 523

QY 69 VPIVNLKDEVLSPSWDSLFGSQGQLQPGARIFSPDGRDLRHPAWPQKSVWHGSDPSGR 128

Db 524 T-VVNVAGHLLFPFSRVFNGA--QMPHAKLFSFDRHDVLNDSRPDKRVWHGSKDGGI 580

QY 129 RLMESYCETWRTTGTATGQASSLLSGLRLEQKAAS--CHNSYIVLCIE 175

Db 581 R-AEQYCDGWRADSSLTSLAGHISNTSIFQSSGSEKCNKLVLCVE 628

RESULT 6

H59432

RhoGAP protein homolog [imported] - human

C/Species: Homo sapiens (man)

C/Date: 03-Jun-2002 #sequence\_revision 03-Jun-2002 #text\_change 23-Sep-2002

C/Accession: H59432

R/Rhodes, S.

submitted to GenBank, October 2001

A/Description: Novel human gene mapping to chromosome 13, similar to rat RhoGAP.

A/Reference number: H59432

A/Accession: H59432

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-995 <RHO>

A/Cross-references: GB:NP\_443083; PID:g16445031; PIDN:NP\_443083.1

Query Match 9.2%; Score 85; DB 2; Length 995;

Best Local Similarity 23.4%; Pred. No. 4.7;

Matches 37; Conservative 22; Mismatches 47; Indels 52; Gaps 7;

QY 52 SRLQDLYSIVRRDRGSVP-----IVNLKDEVLSPSWDSLFGSQGQLQPG 97

Db 36 SRVDDLYTLPRGDRNGSPGGTGMRTTSSSVLTDLSEPEVCISIHSESSGSDRSQPG 95

QY 98 -----ARIFSPD---GRDVLRHPAWPQKSVWHGSDPSGRRLMESYCETWRT-- 140

Db 96 QCCTDNFVMLDAPLVSSSLPQPPRDVLNHPFHPK-----NEKPTRARAKSLKRMETLR 149

QY 141 -----ETTGTATGQASSLLSGLRLEQ-----KAASC 165

Db 150 KGGAHGRHKSGRTG--GLVISGPMLOQPEPESFKAMQC 185

RESULT 7

E97241

hypothetical protein CAC2776 [imported] - Clostridium acetobutylicum

C/Species: Clostridium acetobutylicum

C/Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 14-Sep-2001

C/Accession: E97241

R/Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.

J. Bacteriol. 183, 4823-4838, 2001

A/Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clostridium acetobutylicum

A/Reference number: A96900; MUID:21359325; PMID:21359325

A/Accession: E97241

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-286 <KUR>

A/Cross-references: GB:AE001437; PIDN:AAK80720.1; PID:g15025814; GSPDB:GN00168

A/Experimental source: Clostridium acetobutylicum ATCC824

C/Genetics:

A/Gene: CAC2776

Query Match 8.6%; Score 79.5; DB 2; Length 286;

Best Local Similarity 24.1%; Pred. No. 3.6;

Matches 38; Conservative 21; Mismatches 42; Indels 57; Gaps 8;

QY 24 RGIRGADFQCFQCARAVGLSGTFRFLS-----SRLQDL--YSIVRR-----A 64

Db 42 RGIVVFDISSPIEESI-LSATLKLVSYNYSOLKSIKFSILQKYSINTVYNTQPII 100

QY 65 DRGSVPIVNLKDEV-----LSPSW-----DSLFGSQGQLQPGARI 100

Db 101 DTNYVGITNTNEINEFINVDLSNITSQWHNGSVANLGMVHGDENSGS-----I 151

QY 101 FSPDGRDVLRHPAWPQKSVWHGSDPSGRRLL---MESH 134

Db 152 VGFAGISAINSLWPRNLNVQFSANSSGRVLTITYTMESH 189

RESULT 8

A53798

58K membrane-associated protein - rat

N/Alternate names: 58K microfilament-associated protein

C/Species: Rattus norvegicus (Norway rat)

C/Date: 07-Oct-1994 #sequence\_revision 07-Oct-1994 #text\_change 24-Nov-1999

C/Accession: A53798

R/Juang, S.H.; Huang, J.; Li, Y.; Salas, P.J.I.; Fregien, N.; Carraway, C.A.C.; Carraway, J. Biol. Chem. 269, 15067-15075, 1994

A/Title: Molecular cloning and sequencing of a 58-kDa membrane- and microfilament-associated protein

A/Reference number: A53798; MUID:94253065; PMID:8195143

A/Accession: A53798

A/Status: preliminary; not compared with conceptual translation

A/Molecule type: mRNA

A/Residues: 1-427 <JUA>

A/Cross-references: GB:U15425; GB:S70214; NID:g562083; PIDN:AAB54085.1; PID:g562084

C/Superfamily: mammalian retrovirus gag polyprotein I

C/Keywords: actin binding; monomer; phosphoprotein

Query Match 8.6%; Score 79; DB 2; Length 427;

Best Local Similarity 21.7%; Pred. No. 6.6;

Matches 40; Conservative 37; Mismatches 79; Indels 28; Gaps 8;

QY 12 LVALNTPLSGMREGIR---GADFQCFQCARAVGLSGTFRFLSRLQDLYSIVRR---A 64

Db 185 LEAAPSPVAGRLGRREVAPDSTSQAPFLRQAGGQMYPFSA--DIYNWKQHNPFF 241

QY 65 DRGSVPIVNLKDEVL---SPSWD-----SLFGSQGQ---LQPGARIFSPDGRDVL- 110

Db 242 SKDPVALTNLIESVLLTHQPTWDDIQQLQALLTSEEKQKVLEARKHVLGDNGRPTLLP 301

QY 111 ---HPAWP-QKSVWHGSDPSGRRLMESYCETWRTTGTATGQASSLLSGLRLEQKAASC 166

Db 302 BEIDDAFFLTREDWDFTTAEGRRHLRLYRQLLALLAGLRGAARRPTNLAQVKVQVEAETP 361

QY 167 NSYI 170

Db 362 SAFL 365

RESULT 9

AD0921

probable lipopolysaccharide biosynthesis protein STY3629 [imported] - Salmonella enterica

C/Species: Salmonella enterica subsp. enterica serovar Typhi

A/Note: this species has also been called Salmonella typhi

C/Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002

C/Accession: AD0921

R/Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Gaora, P.

Nature 413, 848-852, 2001

A/Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; A/Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhi

A/Reference number: AB0502; MUID:21534947; PMID:11677608

A/Accession: AD0921

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-416 <PAR>

A/Cross-references: GB:AL513382; PIDN:CAD09390.1; PID:g16504508; GSPDB:GN00176

C/Genetics:

A/Gene: STY3629

Query Match 8.5%; Score 78.5; DB 2; Length 416;

Best Local Similarity 24.4%; Pred. No. 7.1;

Matches 38; Conservative 15; Mismatches 56; Indels 47; Gaps 7;



QY	3	HQDFQPVLHLVAL-----NTPLSGGMRGIRGAD-----FQCQQAR	38
D6	112	HTDYQGRLVLRVALVMGIWANLLALMKGFDRDASGNALSIVGSFLGVAAWYLCT--R	168
QY	39	AVGLSGTFRAFLSSRLQLDLYSIVRR-----DRGSVPINVLDKDEVLPSPWDLSLFGSQQQL	94
D6	169	VGGYEG---ALLGLALVPALVVWPAGIILIKRGTVPLRYLK-----PSWDNGLAGQLSKF	220
QY	95	QPGARIFSF-----GRDVLHRHPAWPKSVWHG	122
D6	221	TLMALITSVTMPVAYVMNRNQLAAHYSWSVDVGIWQG	256
RESULT 10			
S53597			
C:	Species: Euglena gracilis (clone GC18 and others) - Euglena gracilis (var. bacillaris)		
A:	Variety: var. bacillaris		
C:	Date: 01-Mar-1996 #sequence revision 19-Apr-1996 #text_change 09-Sep-1997		
C:	Accession: S53597; S71160; S71481; S19138		
R:	Schwartzbach, S.D.		
submitted to the EMBL Data Library, July 1994			
A:	Reference number: S53597		
A:	Accession: S53597		
A:	Molecule type: DNA		
A:	Residues: 1-1055 <SCH>		
A:	Cross-references: EMBL:X61361; NID:g510387; PID:g510388		
A:	Experimental source: clone GC18; var. bacillaris		
A:	Note: this is a revision to the sequence from reference S19138		
R:	Muchhal, U.S.; Schwartzbach, S.D.		
Nucleic Acids Res.	22, 5737-5744, 1994		
A:	Title: Characterization of the unique intron-exon junctions of Euglena gene(s) encoding A:Reference number: S53596; MUID:95140640; PMID:7838730		
A:	Accession: S71160		
A:	Molecule type: DNA		
A:	Residues: 977-1055 <MUC1>		
A:	Cross-references: EMBL:X61361		
A:	Experimental source: clone GC18; var. bacillaris		
A:	Accession: S71481		
A:	Molecule type: mRNA		
A:	Residues: 977-1055 <MUC2>		
A:	Cross-references: EMBL:X61361		
A:	Experimental source: clone CLH09; clone CLH22; var. bacillaris		
R:	Muchhal, U.S.; Schwartzbach, S.D.		
Plant Mol. Biol.	18, 287-299, 1992		
A:	Title: Characterization of a Euglena gene encoding a polyprotein precursor to the light A:Reference number: S19138; MUID:92119250; PMID:1731990		
A:	Accession: S19138		
A:	Molecule type: DNA		
A:	Residues: 1-374; 376-444, 'A', 446-1043, 'AQQLLF', 1048, 'RFSVTTL', 1055 <MUC>		
A:	Cross-references: EMBL:X61361		
A:	Note: this sequence has been revised in reference S53597		
C:	Genetics:		
A:	Introns: 103/3; 137/3; 375/3; 403/1; 444/2; 598/3; 857/2; 976/3; 1043/3		
Query Match 8.5%; Score 78; DB 2; Length 1055;			
Best Local Similarity 24.2%; Pred. No. 25;			
Matches 37; Conservative 19; Mismatches 63; Indels 34; Gaps 5;			
QY	20	SGGMERGIRGADFQCFOQA---RAVGLSGTFRAPFLSSRLQLDLYS-----IVRAD	65
D6	25	SAGSVGDFFGADLDPLYPGGPDPGLGLADDPEAFLELVKEVKNGRLAMVAIFGSVOGIL	84
QY	66	RGSVPIVNLKDENVLPSPWDLSLFGSQQLQPGAIRIPISFDGRDVLRHPPAWPKSVTHGSDP	125
D6	85	TGKGPFVENWVDHLTDTPFENNLF-----QLTSGFAMLATSGRKAKAAPKSDNLSQWYGPD-	138
QY	126	SGRRLMESYCETWRTEGTG-ATGOASSLLSGRL	157
D6	139	-----RAKWGLPTGEVPSYLTGEL	158

RESULT 11  
T10011  
probable penicillin-binding protein - Mycobacterium leprae  
C/Species: Mycobacterium leprae  
C/Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_cl  
C/Accession: T10011  
R/Cole, S.T.  
submitted to the EMBL Data Library, August 1997  
A/Reference number: Z16916  
A/Accession: T10011  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-474 <COL>  
A/Cross-references: EMBL:Z70722; NID:e1059634; PID:e338505  
C/Genetics:  
A/Gene: pbpA

RESULT 12  
B86911  
probable penicillin-binding protein [imported] - Mycobacterium leprae  
C/Species: Mycobacterium leprae  
C/Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001  
C/Accession: B86911  
R/Cole, S.T.; Eiglmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler  
R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.  
eam, M.A.; Rutherford, K.M.  
Nature 409, 1007-1011, 2001  
A/Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squar  
A/Title: Massive gene decay in the leprosy bacillus.  
A/Reference number: A86909; MUID:21128732; PMID:11234002  
A/Accession: B86911  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-492 <STO>  
A/Cross-references: GB:AL450380; NID:gl3092428; PIDN:CAC29526.1; GSPDB:GN00147  
C/Genetics:  
A/Gene: pbpA

RESULT 13

T30283  
polyketide synthase - Streptomyces sp. (strain MA6548)  
C;Species: Streptomyces sp.  
A;Variety: strain MA6548  
C;Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 17-Nov-2000  
C;Accession: T30283  
R;Motamedi, H.; Cai, S.J.; Shafiee, S.J.; Elliston, K.O.  
Eur. J. Biochem. 244, 74-80, 1997  
A;Title: Structural organization of a multifunctional polyketide synthase involved in the  
A;Reference number: Z20806; MUID:97217427; PMID:9063448  
A;Accession: T30283  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-6420 <MOT>  
A;Cross-references: EMBL:Y10438; NID:e1014806; PID:e290681; PIDN:CAA71463.1  
C;Genetics:  
A;Note: fkbA  
C;Superfamily: 3-oxoacyl-[acyl-carrier-protein] synthase I homology; acyl carrier protein  
C;Keywords: carrier protein  
F;51-433/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>  
F;1930-2325/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>  
F;3391-3462/Domain: acyl carrier protein homology <ACP1>  
F;3505-3900/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS3>  
F;3983-4254/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>  
F;5307-5378/Domain: acyl carrier protein homology <ACP2>  
F;5431-5831/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS4>  
F;5932-6206/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT2>  
F;6293-6364/Domain: acyl carrier protein homology <ACP3>  
  
Query Match 8.4%; Score 77.5; DB 2; Length 6420;  
Best Local Similarity 23.0%; Pred. No. 2.5e+02;  
Matches 47; Conservative 25; Mismatches 73; Indels 59; Gaps 9;  
  
QY 4 QDFQVHLVALNPLSGGMR-----GIRGADFQCFQ-----A 37  
Db 5406 RDQDEPIAIVAMACRLPGGVTSPEBLWRLVESGTDALTAPGDRGWDLDALYDPDPAVG 5465  
  
QY 38 RAVGLSGTFRFLSSRLQDLYSIVRRADRGSVPIVNLKDEVLSFSGSQGLQPG 97  
Db 5466 KAYSILRGGLFGAAEFDAFFDISPESLGMDFPQRL---LLETAWAEI---ERGRINPA 5519  
  
QY 98 ARIFSFDRDVLHRPAWPQKSVVHSGDPSGRRRLMESYCTWRTTGTATGQASSLLSGRL 157  
Db 5520 ----SLHGRI-----GVYVGAAAGYGLGAE-----DTEGNAITGSGTSLLSGRL 5561  
  
QY 158 -----LEOKA-----ASCHNSYIVL 172  
Db 5562 AYVLGLEGPAVTVDACSSSLVAL 5585  
  
RESULT 14  
T45272  
methyltransferase homolog [imported] - Streptomyces coelicolor (A3(2))  
C;Species: Streptomyces coelicolor  
A;Variety: A3(2)  
C;Date: 31-Jan-2000 #sequence\_revision 31-Jan-2000 #text\_change 11-May-2000  
C;Accession: T45272  
R;Martinez-Costa, O.H.; Martin-Triana, A.J.; Martinez, E.; Fernandez-Moreno, M.A.; Malpa  
J. Bacteriol. 181, 4353-4364, 1999  
A;Title: An additional regulatory gene for actinorhodin production in Streptomyces livida  
A;Reference number: Z22953; MUID:99328982; PMID:10400594  
A;Accession: T45272  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-208 <MAR>  
A;Cross-references: EMBL:Y18817; PIDN:CAB51133.1  
A;Experimental source: A3(2); strain J1501  
C;Genetics:  
A;Note: ORF8

Query Match 8.3%; Score 77; DB 2; Length 208;  
Best Local Similarity 25.6%; Pred. No. 4.3;  
Matches 40; Conservative 20; Mismatches 52; Indels 44; Gaps 7;

QY 39 AVGLSGTFRFLSSR-----LQDLYSIVRRADRGSV-----IVNLKDEVLSPSWD----- 84  
Db 64 AVDISGVAERLAGHARTHGLGLVDVAVRHDLRDSFPEGFDLVSA--HYLHTFPDLDR 121  
QY 85 SLFSGSQGLQPGARI-----FSFDGRDVLHRPAWPQKSVVHSGDPSGRRRLMESY 134  
Db 122 SVLRGAHAHLRPGGRLLVVDHGSTAPWSWDQDPDARHPAPQEVAAADLALDPA----- 173  
QY 135 CETWRT-----TTGATGQASSLLSGRLLEOKA 162  
Db 174 --TWRVERAEAPRRATATGPGGRTAEVVDHVLVRRRA 207

RESULT 15  
B89781  
conserved hypothetical protein SA0184 [imported] - Staphylococcus aureus (strain N315)  
C;Species: Staphylococcus aureus  
C;Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 22-Oct-2001  
C;Accession: B89781  
R;Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc  
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;  
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.  
Lancet 357, 1225-1240, 2001  
A;Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.  
A;Reference number: A89758; MUID:21311952; PMID:11418146  
A;Accession: B89781  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-351 <KUR>  
A;Cross-references: GB:BA000018; PID:g13700106; PIDN:BA041405.1; GSPDB:GN00149  
A;Experimental source: strain N315  
C;Genetics:  
A;Gene: SA0184

Query Match 8.3%; Score 77; DB 2; Length 351;  
Best Local Similarity 23.7%; Pred. No. 8.1;  
Matches 41; Conservative 20; Mismatches 62; Indels 50; Gaps 8;  
  
QY 23 MRGIRGADFQCFQQAQAVGLSGTFRFLSSRLQDL-----YSIVRRADRG-SVPIV 72  
Db 97 IEAIMAQGLKCCCLNASIIS-----RELLTSLHQQLNDFLLSFCHNYYPRTDGLSVDLV 151  
  
QY 73 NLKDEVLSPSWDSLSFSGSQGLQPGARIFSFDRDVLR---HPAWPQKSVVHSGDP--SG 127  
Db 152 NKKNELIY-----QFNPKAQIYGFIVGSLRGLRGLHKLGLPTIEATRHSHPVVAA 199  
  
QY 128 RRLMESYCTWRTTGTATGQASSLLSGRLLEOKAAS-----CHNSYIVLCIE 175  
Db 200 KLLQE-----TGVSEVLVGDLSLIEMRQAKQLIDFCKHRHFTLCIE 239

Search completed: February 17, 2004, 09:58:26  
Job time : 24 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 17, 2004, 09:46:12 ; Search time 17 Seconds  
(without alignments)  
484.098 Million cell updates/sec

Title: US-09-589-777C-2\_COPY\_1\_175

Perfect score: 923

Sequence: 1 HTHQDFQPVHLVALNTPLS.....RLLEQKAASCHNSYIVLCIE 175

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_41:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	923	100.0	1527	1 CAIH_MOUSE	P39061 mus musculu
2	813	88.1	1516	1 CAIH_HUMAN	P39060 homo sapien
3	535	58.0	1388	1 CAIE_HUMAN	P39059 homo sapien
4	85	9.2	995	1 SR13_HUMAN	Q9y3m8 homo sapien
5	77	8.3	1574	1 SYJ1_RAT	Q62910 rattus norv
6	76.5	8.3	448	1 HGD_BRAJA	Q89xh1 bradyrhizob
7	76.5	8.3	1050	1 ULK1_HUMAN	O75385 homo sapien
8	76	8.2	309	1 YHCC_ECOLI	P45476 escherichia
9	76	8.2	1575	1 SYJ1_HUMAN	O43426 homo sapien
10	75	8.1	505	1 CATA_METBA	O93662 methanosarc
11	75	8.1	1324	1 SYJ1_BOVIN	O18964 bos taurus
12	75	8.1	7073	1 RIAB_CVHSA	P59641 h replicase
13	74.5	8.1	611	1 PHBC_RHIME	P50176 r poly-beta
14	72.5	7.9	419	1 P47K_PSECL	P31521 pseudomonas
15	72.5	7.9	739	1 VNUC_EBOG4	Q9qce9 ebola virus
16	72.5	7.9	800	1 P5CS_CABEL	P54889 c probable
17	72	7.8	530	1 MATP_MOUSE	P58355 mus musculu
18	72	7.8	1233	1 NME3_HUMAN	Q14957 homo sapien
19	72	7.8	1955	1 AGRI_CHICK	P31696 gallus gall
20	71	7.7	1289	1 CSAB_BACUD	Q45753 bacillus th
21	70.5	7.6	309	1 YEEY_ECOLI	P76369 escherichia
22	70.5	7.6	319	1 HALL_RAT	P15978 rattus norv
23	70	7.6	306	1 PERE_ARMRU	P59121 armoracia r
24	69.5	7.5	437	1 ADFF_HUMAN	Q99541 homo sapien
25	69.5	7.5	692	1 GYRB_BARBA	P94281 bartonella
26	69.5	7.5	738	1 VNUC_EBOB	Q9qp77 ebola virus
27	69.5	7.5	739	1 VNUC_EBOZ5	O72142 ebola virus
28	69.5	7.5	739	1 VNUC_EBOZM	P18272 ebola virus
29	69.5	7.5	860	1 SZ1A_BRARE	Q9w7j1 brachydanio
30	69	7.5	174	1 CARQ_MYXXA	Q06909 myxococcus
31	69	7.5	692	1 VNUC_MABVM	P27588 marburg vir
32	69	7.5	695	1 VNUC_MABVP	P35263 marburg vir
33	69	7.5	1023	1 DPOL_ADEB3	O72540 bovine aden

34	69	7.5	1357	1 YJ03_YEAST	P47104 saccharomyc
35	69	7.5	5217	1 HTS1_COCCA	Q01886 cochliobolu
36	68.5	7.4	266	1 NANH_BACFR	P31206 bacteroides
37	68.5	7.4	1248	1 SYJ2_RAT	O55207 rattus norv
38	68	7.4	362	1 SERC_BACCI	Q59196 bacillus ci
39	68	7.4	587	1 HEMO_HUMAN	P22557 homo sapien
40	68	7.4	735	1 RN5A_MOUSE	Q05921 mus musculu
41	68	7.4	1051	1 ULK1_MOUSE	O70405 mus musculu
42	68	7.4	1203	1 RPA2_YEAST	P22138 saccharomyc
43	68	7.4	1220	1 C5AC_BACTU	P56955 bacillus th
44	68	7.4	1385	1 C5AA_BACUD	Q45760 bacillus th
45	68	7.4	1443	1 SYJ2_HUMAN	O15056 homo sapien

ALIGNMENTS

RESULT 1  
CAIH\_MOUSE  
ID CAIH\_MOUSE STANDARD; PRT; 1527 AA.  
AC P39061; Q61437; Q62002;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Collagen alpha 1(XVII) chain precursor [Contains: Endostatin].  
GN COL18A1.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORM SHORT).  
RC STRAIN=BALB/c; TISSUE=Liver;  
RX MEDLINE=94245707; PubMed=8188673;  
RA Rehn M.V., Hintikka E., Pihlajaniemi T.;  
RT "Primary structure of the alpha 1 chain of mouse type XVIII collagen,  
RT partial structure of the corresponding gene, and comparison of the  
RT alpha 1(XVIII) chain with its homologue, the alpha 1(XV) collagen  
RT chain.";  
RL J. Biol. Chem. 269:13929-13935(1994).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORMS LONG AND SHORT).  
RA Rehn M., Hintikka E., Pihlajaniemi T.;  
RT "Characterization of the mouse gene for the alpha-1 chain of type  
RT XVIII collagen (COL18A1) reveals that the three variant N-terminal  
RT polypeptide forms are transcribed from two widely separated  
RT promoters.";  
RL Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE OF 213-1140 FROM N.A. (ISOFORM SHORT).  
RX MEDLINE=94240112; PubMed=8183894;  
RA Rehn M.V., Pihlajaniemi T.;  
RT "Alpha 1(XVIII), a collagen chain with frequent interruptions in the  
RT collagenous sequence, a distinct tissue distribution, and homology  
RT with type XV collagen.";  
RL Proc. Natl. Acad. Sci. U.S.A. 91:4234-4238(1994).  
RN [4]  
RP SEQUENCE OF 240-1527 FROM N.A.  
RC TISSUE=Liver;  
RX MEDLINE=94240111; PubMed=8183893;  
RA Oh S.P., Kamagata Y., Muragaki Y., Timmons S., Ooshima A., Olsen B.R.;  
RT "Isolation and sequencing of cDNAs for proteins with multiple domains  
RT of Gly-Xaa-Yaa repeats identify a distinct family of collagenous  
RT proteins.";  
RL Proc. Natl. Acad. Sci. U.S.A. 91:4229-4233(1994).  
RN [5]  
RP CHARACTERIZATION OF ENDOSTATIN, AND PARTIAL SEQUENCE.  
RX MEDLINE=97160848; PubMed=9008168;  
RA O'Reilly M.S., Boehm T., Shing Y., Fukai N., Vasios G., Lane W.S.,  
RA Flynn E., Birkhead J.R., Olsen B.R., Folkman J.;  
RT "Endostatin: an endogenous inhibitor of angiogenesis and tumor  
RT growth.";  
RL Cell 88:277-285(1997).



Query Match 100.0%; Score 923; DB 1; Length 1527;  
Best Local Similarity 100.0%; Pred. No. 1.3e-81;  
Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	HTHQDFQPVHLVALNTPLSGGMRGIRGADFQCFQQAARAVGLSGTFRFLSSRLQDLYSI	60
DB	1344	HTHQDFQPVHLVALNTPLSGGMRGIRGADFQCFQQAARAVGLSGTFRFLSSRLQDLYSI	1403
QY	61	VRRADRGSVPIVNLKDEVLSFSGSQQLQPGARIFSFDRDVLRRHPAPQKSVW	120
DB	1404	VRRADRGSVPIVNLKDEVLSFSGSQQLQPGARIFSFDRDVLRRHPAPQKSVW	1463
QY	121	HGSDPSGRRLMESYCETWRTTGTGATGCAASSLLSGRLLEQKAASCHNSYIVLCIE	175
DB	1464	HGSDPSGRRLMESYCETWRTTGTGATGCAASSLLSGRLLEQKAASCHNSYIVLCIE	1518

RESULT 2

CA1H_HUMAN	STANDARD;	PRT;	1516 AA.
ID_CALH_HUMAN	Q9UK38; Q9Y6Q7; Q9Y6Q8;		
AC	P39060; Q9UK38; Q9Y6Q7; Q9Y6Q8;		
DT	01-FEB-1995 (Rel. 31, Created)		
DT	16-OCT-2001 (Rel. 40, Last sequence update)		
DT	15-SEP-2003 (Rel. 42, Last annotation update)		
DE	Collagen alpha 1(XVII) chain precursor [Contains: Endostatin].		
GN	COL18A1.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Euthera; Primates; Catarrhini; Hominiidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=98164096; PubMed=9503365;		
RA	Saarela J., Ylikarppa R., Rehn M., Purmonen S., Pihlajaniemi T.;		
RT	"Complete primary structure of two variant forms of human type XVIII		
RT	collagen and tissue-specific differences in the expression of the		
RT	corresponding transcripts.";		
RL	Matrix Biol. 16:319-328(1998).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=20289799; PubMed=10830953;		
RA	Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,		
RA	Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Groner Y.,		
RA	Soeda E., Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K.,		
RA	Pollay A., Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,		
RA	Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,		
RA	Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,		
RA	Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,		
RA	Minoshima S., Shimizu N., Nordsiek G., Hornischer K., Brandt P.,		
RA	Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloeker H.,		
RA	Ramser J., Beck A., Klages S., Hennig S., Riesselmann L., Dagand E.,		
RA	Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,		
RA	Lehrach H., Reinhardt R., Yaspo M.-L.;		
RT	"The DNA sequence of human chromosome 21.";		
RL	Nature 405:311-319(2000).		
RN	[3]		
RP	SEQUENCE OF 834-1516 FROM N.A.		
RX	MEDLINE=94245237; PubMed=8188291;		
RA	Oh S.P., Warman M.L., Seldin M.F., Cheng S., Knoll J.H., Timmons S.,		
RA	Olsen B.R.;		
RT	"Cloning of cDNA and genomic DNA encoding human type XVIII collagen		
RT	and localization of the alpha 1(XVII) collagen gene to mouse		
RT	chromosome 10 and human chromosome 21.";		
RL	Genomics 19:494-499(1994).		
RN	[4]		
RP	SEQUENCE OF 1334-1516 FROM N.A.		
RC	TISSUE=Placenta;		
RA	Zhi-Yong H., Biao L., Wei-Jie Z., Xiang-Fu W.;		
RT	"Cloning and expression of human endostatin gene in Escherichia		
RT	coli.";		
RL	Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.		
RN	[5]		

RP	INVOLVEMENT IN KNOBLOCH SYNDROME.
RX	MEDLINE=20400145; PubMed=10942434;
RA	Sertie A.L., Sossi V., Camargo A.A., Zatz M., Brahe C.,
RA	Passos-Bueno M.R.;
RT	"Collagen XVIII, containing an endogenous inhibitor of angiogenesis
RT	and tumor growth, plays a critical role in the maintenance of retinal
RT	structure and in neural tube closure.";
RL	Hum. Mol. Genet. 9:2051-2058(2000).
RN	[6]
RP	VARIANT ASN-1437.
RX	MEDLINE=21518361; PubMed=11606364;
RA	Iughetti P., Suzuki O., Godoi P.H., Alves V.A., Sertie A.L.,
RA	Zorick T., Soares F., Camargo A., Moreira E.S., di Loreto C.,
RA	Moreira-Filho C.A., Simpson A., Oliva G., Passos-Bueno M.R.;
RT	"A polymorphism in endostatin, an angiogenesis inhibitor, predisposes
RT	for the development of prostatic adenocarcinoma.";
RL	Cancer Res. 61:7375-7378(2001).
CC	-!- FUNCTION: COL18A1 PROBABLY PLAYS A MAJOR ROLE IN DETERMINING THE
CC	RETINAL STRUCTURE AS WELL AS IN THE CLOSURE OF THE NEURAL TUBE.
CC	-!- FUNCTION: ENDOSTATIN POTENTLY INHIBITS ENDOTHELIAL CELL
CC	PROLIFERATION AND ANGIOGENESIS. MAY INHIBIT ANGIOGENESIS BY
CC	BINDING TO THE HEPARAN SULPHATE PROTEOGLYCANS INVOLVED IN GROWTH
CC	FACTOR SIGNALLING.
CC	-!- ALTERNATIVE PRODUCTS:
CC	Event=Alternative splicing; Named isoforms=2;
CC	Name=Long; Synonyms=NC-493;
CC	Isoid=P39060-1; Sequence=Displayed;
CC	Name=Short; Synonyms=NC1-303;
CC	Isoid=P39060-2; Sequence=VSP 001155, VSP 001156;
CC	-!- TISSUE SPECIFICITY: PRESENT IN MULTIPLE ORGANS WITH HIGHEST LEVELS
CC	IN LIVER, LUNG AND KIDNEY.
CC	-!- PTM: Prolines at the third position of the tripeptide repeating
CC	unit (G-X-Y) are hydroxylated in some or all of the chains.
CC	-!- POLYMORPHISM: There is an association between a polymorphism in
CC	position 1437 and prostate cancer. Heterozygous Asn-1437
CC	individuals have a 2.5 times increased chance of developing
CC	prostate cancer as compared with homozygous Asp-1437 individuals.
CC	-!- DISEASE: Defects in COL18A1 are a cause of Knobloch syndrome (KNO)
CC	[MIM:267750]; an autosomal recessive disorder defined by the
CC	occurrence of high myopia, vitreoretinal degeneration with retinal
CC	detachment, macular abnormalities and occipital encephalocele.
CC	-!- SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH
CC	INTERRUPTED HELICES (FACIT) FAMILY.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).
CC	-----
DR	EMBL; AF018081; AAC39658.1; -.
DR	EMBL; AF018082; AAC39659.1; -.
DR	EMBL; AL163302; CAB90482.1; -.
DR	EMBL; L22548; AAA51864.1; -.
DR	EMBL; AF184060; AAF01310.1; ALT_INIT.
DR	PDB; 1BNL; 02-DEC-98.
DR	GlycoSuiteDB; P39060; -.
DR	Genew; HGNC:2195; COL18A1.
DR	MIM; 120328; -.
DR	MIM; 267750; -.
DR	GO; GO:0005581; C:collagen; TAS.
DR	GO; GO:0008181; F:tumor suppressor; TAS.
DR	GO; GO:0007397; P:histogenesis and organogenesis; TAS.
DR	GO; GO:0008285; P:negative regulation of cell proliferation; TAS.
DR	GO; GO:0007048; P:oncogenesis; TAS.
DR	GO; GO:0007601; P:vision; TAS.
DR	InterPro; IPR000087; Collagen.
DR	InterPro; IPR001791; Laminin_G.
DR	InterPro; IPR003129; TSPN.
DR	Pfam; PF01391; Collagen; 7.
DR	Pfam; PF02210; TSPN; 1.







DR MIM; 120325; --  
DR InterPro; IPR000087; Collagen.  
DR InterPro; IPR001791; Laminin\_G.  
DR InterPro; IPR003129; TSPN.  
DR Pfam; PF01391; Collagen; 5.  
DR Pfam; PF02210; TSPN; 1.  
DR SMART; SM00282; LamG; 1.  
DR SMART; SM00210; TSPN; 1.  
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
KW Cell adhesion; Collagen; Glycoprotein; Signal.  
FT SIGNAL 1 25  
FT CHAIN 26 1388  
FT DOMAIN 26 555  
FT DOMAIN 556 573  
FT DOMAIN 574 618  
FT DOMAIN 619 732  
FT DOMAIN 733 763  
FT DOMAIN 764 798  
FT DOMAIN 799 822  
FT DOMAIN 823 867  
FT DOMAIN 868 878  
FT DOMAIN 879 949  
FT DOMAIN 950 983  
FT DOMAIN 984 1013  
FT DOMAIN 1014 1027  
FT DOMAIN 1028 1045  
FT DOMAIN 1046 1052  
FT DOMAIN 1053 1107  
FT DOMAIN 1108 1117  
FT DOMAIN 1118 1132  
FT DOMAIN 1133 1388  
FT DOMAIN 358 555  
FT REPEAT 358 408  
FT REPEAT 409 459  
FT REPEAT 460 509  
FT REPEAT 510 555  
FT CARBOHYD 306 306  
FT CARBOHYD 324 324  
FT CARBOHYD 687 687  
FT CARBOHYD 807 807  
FT CARBOHYD 814 814  
FT CARBOHYD 1046 1046  
FT CONFLICT 10 10  
FT CONFLICT 49 49  
FT CONFLICT 95 95  
FT CONFLICT 150 150  
FT CONFLICT 204 204  
FT CONFLICT 409 409  
SQ SEQUENCE 1388 AA; 141930 MW; 60822AD925A3093D CRC64;  
Query Match 58.0%; Score 535; DB 1; Length 1388;  
Best Local Similarity 60.4%; Pred. No. 6.1e-44;  
Matches 102; Conservative 23; Mismatches 40; Indels 4; Gaps 1;  
QY 7 QPVLHLVALNTPLSGGMRGIRGADFCQQAQAVGLSGTFFAFLSSRLQDLYSIVRRADR 66  
Db 1215 KPALHLAALNMPFSGDIR---ADFQCFKQARAAGLLSTYRAFLSSHLQDLSTIVRKAER 1270  
QY 67 GSVPIVNLKDEVLSFSGSQGLQPGARIFSFQDGRDVLHPAPWPKSVWHGSDPS 126  
Db 1271 YSLPIVNLKGVLFNNWDSIFSGHGGQFMNHIPIYSFDGRDINTDPSWPQKVIWHGSSPH 1330  
QY 127 GRRLMESYCETWRTTGTATGQASSLLSGLLLEQKAASCHNSYIVLCIE 175  
Db 1331 GVRLVDNYCEAWRTADTAVTGLASPLSTGKILDQKAYSCANRLIVLCIE 1379  
RESULT 4  
SR13\_HUMAN STANDARD; PRT; 995 AA.  
ID SR13\_HUMAN  
AC Q9Y3M8;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE STAR-related lipid transfer protein 13 (StARD13) (START domain-  
DE containing protein 13) (46H23.2).  
GN STARD13 OR GT650.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Rhodes S.;  
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Hunt A.;  
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: May function as a GTPase-activating protein.  
CC -!- SIMILARITY: Contains 1 Rho-GAP domain.  
CC -!- SIMILARITY: Contains 1 START domain.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AL049801; CAB42562.1; --  
DR EMBL; Z84483; CAC94774.1; --  
DR PIR; H59432; H59432.  
DR Genew; HGNC:19164; STARD13.  
DR InterPro; IPR000198; RhoGAP.  
DR InterPro; IPR002913; START.  
DR Pfam; PF00620; RhoGAP; 1.  
DR Pfam; PF01852; START; 1.  
DR SMART; SM00324; RhoGAP; 1.  
DR SMART; SM00234; START; 1.  
DR PROSITE; PS50238; RHO-GAP; 1.  
DR PROSITE; PS50848; START; 1.  
KW GTPase activation.  
FT DOMAIN 545 750 RHO-GAP.  
FT DOMAIN 781 989 START.  
SQ SEQUENCE 995 AA; 111191 MW; 3F608FA94A4EF8BF CRC64;  
Query Match 9.2%; Score 85; DB 1; Length 995;  
Best Local Similarity 23.4%; Pred. No. 2.4;  
Matches 37; Conservative 22; Mismatches 47; Indels 52; Gaps 7;  
QY 52 SRLQDLYSIVRRADRGVSP-----IVNLKDEVLSFSGSQGLQPG 97  
Db 36 SRVDDLYTLPRGDRKNGSPGGTGMRTTSSSVLTDLSEPEVCSIHSESSGGSDRSQPG 95  
QY 98 -----ARIFSF-----GRDVLHPAPWPKSVWHGSDPSGRRRLMESYCETWRT-- 140  
Db 96 QCCTDNPVMDAPLVSSSLPQPPRDVLNHPFHPK-----NEKPTRARAKSFLKRMETLR 149  
QY 141 -----ETTGTATGQASSLLSGLLLEQ-----KAASC 165  
Db 150 KGKANGRHKSGRGIG--GLVISGPMLOQEPESFKAMQC 185  
RESULT 5  
SYJ1\_RAT STANDARD; PRT; 1574 AA.  
ID SYJ1\_RAT  
AC Q62910; O89092; Q62911;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Synaptojanin 1 (EC 3.1.3.36) (Synaptic inositol-1,4,5-trisphosphate 5-  
DE phosphate 1).  
GN SYNJ1.  
OS Rattus norvegicus (Rat).



DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Homogentisate 1,2-dioxygenase (EC 1.13.11.5) (Homogentisicase)  
DE (Homogentisate oxygenase) (Homogentisic acid oxidase).  
GN HMGA OR BLL0343.  
OS Bradyrhizobium japonicum.  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
OC Bradyrhizobiaceae; Bradyrhizobium.  
OX NCBI\_TaxID=375;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=USDA 110;  
RX MEDLINE=22484998; PubMed=12597275;  
RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,  
RA Sasamoto S., Watanabe A., Idesawa K., Iriguchi M., Kawashima K.,  
RA Kohara M., Matsumoto M., Shimo S., Tsuruoka H., Wada T., Yamada M.,  
RA Tabata S.;  
RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium  
Bradyrhizobium japonicum USDA110.";  
RL DNA Res. 9:189-197(2002).  
CC -|- CATALYTIC ACTIVITY: Homogentisate + O(2) = 4-maleylacetate.  
CC -|- COFACTOR: Iron (By similarity).  
CC -|- PATHWAY: Catabolism of tyrosine; third step.  
CC -|- PATHWAY: Catabolism of phenylalanine; fourth step.  
CC -|- PATHWAY: Catabolism of phenylalanine; fourth step.  
CC -|- SIMILARITY: Belongs to the homogentisate dioxygenase family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AP005936; BAC45608.1; ALT\_INIT.  
DR HAMAP; MF 00334; -; 1.  
KW Oxidoreductase; Dioxygenase; Metal-binding; Iron;  
KW Phenylalanine catabolism; Tyrosine catabolism; Complete proteome.  
FT METAL 346 346 IRON (BY SIMILARITY).  
FT METAL 352 352 IRON (BY SIMILARITY).  
FT METAL 382 382 IRON (BY SIMILARITY).  
SQ SEQUENCE 448 AA; 49780 MW; F6DD465E68735D3C CRC64;  
Query Match 8.3%; Score 76.5; DB 1; Length 448;  
Best Local Similarity 25.0%; Pred. No. 6;  
Matches 32; Conservative 22; Mismatches 51; Indels 23; Gaps 7;  
QY 25 GIRGADFCQQAQAVGLSGTFR-----FLSSRLQDLYSIVRRADRGVSP-IVNLKDE 77  
DB 320 GTANIDFVTFPE-RWVADNTFRPWYHNMIMSEFMGLYGVYDAKPGQFVPGMSLHNC 378  
QY 78 VL--SPSWDSLFGSQQLQP-----GARIFSDGRDVLRLHPAWPQKSVVHGDPSGRRLLM 131  
DB 379 MLPHGPDRAFEHASNGELKPKVLGTWAFMFETR-----YPQVTAHAANAS--TLQ 429  
QY 132 ESYCETWR 139  
DB 430 DDYADCWK 437  
RESULT 7  
ULK1\_HUMAN  
ID ULK1\_HUMAN STANDARD; PRT; 1050 AA.  
AC O75385;  
DT 15-JUL-1999 (Rel. 38, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Serine/threonine-protein kinase ULK1 (EC 2.7.1.1-) (Unc-51-like kinase  
DE 1).  
GN ULK1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;

RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98360094; PubMed=9693035;  
RA Kuroyanagi H., Yan J., Seki N., Yamanouchi Y., Suzuki Y.-I.,  
RA Takano T., Muramatsu M.-A., Shirasawa T.;  
RT "Human ULK1, a novel serine/threonine kinase related to UNC-51 kinase  
of Caenorhabditis elegans: cDNA cloning, expression, and chromosomal  
assignment.";  
RL Genomics 51:76-85(1998).  
RN [2]  
RP INTERACTION WITH GABARAP AND GABARAPL2.  
RX MEDLINE=21066693; PubMed=11146101;  
RA Okazaki N., Yan J., Yuasa S., Ueno T., Kominami E., Masuho Y.,  
RA Koga H., Muramatsu M.-A.;  
RT "Interaction of the Unc-51-like kinase and microtubule-associated  
protein light chain 3 related proteins in the brain: possible role of  
vesicular transport in axonal elongation.";  
RL Brain Res. Mol. Brain Res. 85:1-12(2000).  
CC -|- SUBUNIT: Interacts with GABARAP and GABARAPL2.  
CC -|- TISSUE SPECIFICITY: Ubiquitously expressed. Detected in the  
CC following adult tissues: skeletal muscle, heart, pancreas, brain,  
CC placenta, liver, kidney, and lung.  
CC -|- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
CC APO1/UNC-51/ULK1 SUBFAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AF045458; AAC32326.1; -;  
DR HSSP; P24941; 1CKP.  
DR Genew; HGNC:12558; ULK1.  
DR MIM; 603168; -;  
DR GO; GO:0004674; F:protein serine/threonine kinase activity; TAS.  
DR GO; GO:0006468; P:protein amino acid phosphorylation; TAS.  
DR GO; GO:0007165; P:signal transduction; TAS.  
DR InterPro; IPR000719; Prot\_kinase.  
DR InterPro; IPR002290; Ser\_thr\_kinase.  
DR InterPro; IPR001245; Tyr\_kinase.  
DR Pfam; PF00069; pkinase; 1.  
DR PRINTS; PR00109; TYRKINASE.  
DR ProDom; PD000001; Prot\_kinase; 1.  
DR SMART; SM00220; S\_TKc; 1.  
DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
KW Transferase; Serine/threonine-protein kinase; ATP-binding.  
FT DOMAIN 16 278 PROTEIN KINASE.  
FT NP\_BIND 22 30 ATP (BY SIMILARITY).  
FT DOMAIN 287 416 INTERACTION WITH GABARAP AND GABARAPL2.  
FT BINDING 46 46 ATP (BY SIMILARITY).  
FT ACT\_SITE 138 138 BY SIMILARITY.  
FT DOMAIN 297 310 POLY-SER.  
SQ SEQUENCE 1050 AA; 112601 MW; 4ED9B94028E3C138 CRC64;  
Query Match 8.3%; Score 76.5; DB 1; Length 1050;  
Best Local Similarity 30.3%; Pred. No. 17;  
Matches 43; Conservative 13; Mismatches 47; Indels 39; Gaps 9;  
QY 18 PLSGGMRGIRGADFCQQAQAVGLSGTF-RAFLSSRLQDLYSIVRRA-----DRGSVP 70  
DB 674 PLGPGLR--PGED-----PKGFFGRSFSTRLTDL--LLKAAFGTQAPDPGST 718  
QY 71 IVNLKDEVLSFSGQQLQPGARIFSDGRDVLRLHPAWPQKSVVH-GSDPSGR 129  
DB 719 SLQEKPMETAPS-----AGFGGSLHPPGARAGGTSS-----PSPVFTVGSPPSGST 764  
QY 130 LMESYCETWRTTETGATGQASS 151

Db 765 PPQG--PRTRMFSAGPTGSASS 784

RESULT 8

YHCC\_ECOLI STANDARD; PRT; 309 AA.

AC P45476;

DT 01-NOV-1995 (Rel. 32, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Hypothetical protein yhcC.

GN YHCC OR B3211 OR Z4575 OR ECS4090.

OS Escherichia coli, and

OS Escherichia coli O157:H7.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Escherichia.

OX NCBI\_TaxID=562, 83334;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=K12 / MG1655;

RX MEDLINE=97426617; PubMed=9278503;

RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,

RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,

RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,

RA Mau B., Shao Y.;

RT "The complete genome sequence of Escherichia coli K-12.";

RL Science 277:1453-1474(1997).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=O157:H7 / EDL933 / ATCC 700927;

RX MEDLINE=21074935; PubMed=11206551;

RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,

RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,

RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,

RA Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamouisis K.,

RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,

RA Welch R.A., Blattner F.R.;

RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";

RL Nature 409:529-533(2001).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=O157:H7 / RIMD 0509952;

RX MEDLINE=21156231; PubMed=11258796;

RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,

RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,

RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,

RA Kuhara S., Shiba T., Hattori M., Shingawa H.;

RT "Complete genome sequence of enterohaemorrhagic Escherichia coli

RT O157:H7 and genomic comparison with a laboratory strain K-12.";

RL DNA Res. 8:11-22(2001).

RN [4]

RP SEQUENCE OF 1-39 FROM N.A.

RA Ernstring B.R., Denninger J., Blumenthal R.M., Matthews R.G.;

RL Submitted (JUN-1993) to the EMBL/GenBank/DBJ databases.

CC -!- SIMILARITY: TO M.JANNASCHII MJ0486.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL; U18997; AAA58013.1; ALT\_SEQ.

DR EMBL; AE000400; AAC76243.1; -.

DR EMBL; AE005549; AAG58345.1; -.

DR EMBL; AF002564; BAB37513.1; -.

DR EMBL; L20253; -; NOT\_ANNOTATED\_CDS.

DR PIR; B91140; B91140.

DR PIR; E65112; E65112.

DR PIR; E85985; E85985.

DR EcoGene; EGI2809; yhcC.

DR InterPro; IPR005911; Cons\_hypoth1212.

DR InterPro; IPR006638; Elp3.

DR Pfam; PF04055; Radical\_SAM; 1.

DR SMART; SM00729; Elp3; 1.

DR TIGRFAMS; TIGR01212; TIGR01212; 1.

KW Hypothetical protein; Complete proteome.

FT CONFLICT 24 25 KL -> NV (IN REF. 4).

SQ SEQUENCE 309 AA; 34606 MW; 61B3187BB77CA1A9 CRC64;

Query Match 8.2%; Score 76; DB 1; Length 309;

Best Local Similarity 26.1%; Pred. No. 4.3;

Matches 35; Conservative 23; Mismatches 42; Indels 34; Gaps 10;

QY 27 RGADFQCFQQ-----AR-----AVGLSGTFRFLSSRLQDLYSIVRRADRGSV--P 70

Db 159 RGHDFACYQRTTQLARQRLKVCCHLIIVGLPGEQA---ECLQTLRWVETGVGDIKLHP 215

QY 71 IVNLKDEVLSFSGSQQLQPGARI--PSFDGRDVLRLHPAWPQKSVWHGSDPSGR 128

Db 216 LHIVKGSIMAKAWEA-----GRLN-GIELEDYTLTAGEMIRHT--PPEVIYHRISASAR 266

QY 129 R---LMESYCET-W 138

Db 267 RPTLLAPLWCENRW 280

RESULT 9

SYJ1\_HUMAN STANDARD; PRT; 1575 AA.

ID SYJ1\_HUMAN

AC O43426; O43425;

DT 30-MAY-2000 (Rel. 39, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Synaptotagmin 1 (EC 3.1.3.36) (Synaptic inositol-1,4,5-trisphosphate 5-phosphatase 1).

GN SYNJ1.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).

RC TISSUE=Cerebellum;

RX MEDLINE=98088905; PubMed=9428629;

RA Haffner C., Takei K., Chen H., Ringstad N., Hudson A., Butler M.H.,

RA Salcini A.E., Di Fiore P.P., De Camilli P.;

RT "Synaptotagmin 1: localization on coated endocytic intermediates in

RT nerve terminals and interaction of its 170 kDa isoform with Eps15.";

RL FEBS Lett. 419:175-180(1997).

CC -!- FUNCTION: Inositol 5-phosphatase which has a role in clathrin-

CC mediated endocytosis.

CC -!- CATALYTIC ACTIVITY: 1-phosphatidyl-1D-myo-inositol 4,5-

CC bisphosphate + H(2)O = 1-phosphatidyl-1D-myo-inositol 4-phosphate

CC + phosphate.

CC -!- ALTERNATIVE PRODUCTS:

CC Event=Alternative splicing; Named isoforms=2;

CC Name=1; Synonyms=Synaptotagmin-170;

CC IsoId=O43426-1; Sequence=Displayed;

CC Name=2; Synonyms=Synaptotagmin-145;

CC IsoId=O43426-2; Sequence=VSP\_002682, VSP\_002683;

CC -!- TISSUE SPECIFICITY: Concentrated at clathrin-coated endocytic

CC intermediates in nerve terminals. Isoform 1 is more enriched than

CC isoform 2 in developing brain as well as non-neuronal cells.

CC Isoform 2 is very abundant in nerve terminals.

CC -!- DOMAIN: BINDS TO EPS15 (A CLATHRIN COAT-ASSOCIATED PROTEIN) VIA A

CC C-TERMINAL DOMAIN CONTAINING THREE ASN-PRO-PHE (NPF) REPEATS.

CC -!- DOMAIN: THE C-TERMINAL PROLINE-RICH REGION MEDIATES BINDING TO A

CC VARIETY OF SH3 DOMAIN-CONTAINING PROTEINS INCLUDING AMPHIPHYSIN,

CC SH3P4 AND GRB2.

CC -!- SIMILARITY: IN THE CENTRAL SECTION; BELONGS TO THE INOSITOL-1,4,5-

CC TRISPHOSPHATE 5-PHOSPHATASE FAMILY.

CC -!- SIMILARITY: Contains 1 SAC domain.

CC -!- SIMILARITY: Contains 1 RNA recognition motif (RRM) domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AF009040; AAC51922.1; --  
DR EMBL; AF009039; AAC51921.1; --  
DR Genew; HGNC:11503; SYNJ1.  
DR MIM; 604297; --  
DR GO; GO:0005803; C:secretory vesicle; TAS.  
DR GO; GO:0004445; F:inositol-1,4,5-trisphosphate 5-phosphatase . . . ; TAS.  
DR GO; GO:0006899; P:non-selective vesicle transport; TAS.  
DR GO; GO:0006796; P:phosphate metabolism; TAS.  
DR GO; GO:0008099; P:synaptic vesicle endocytosis; TAS.  
DR InterPro; IPR005135; Exo endo\_phos.  
DR InterPro; IPR000300; IPPC.  
DR InterPro; IPR000504; RNA\_rec\_mot.  
DR InterPro; IPR002013; Syn\_N.  
DR Pfam; PF03372; Exo endo\_phos; 1.  
DR Pfam; PF02383; Syn\_N; 1.  
DR SMART; SM00128; IPPC; 1.  
DR PROSITE; PS50102; RRM; 1.  
DR PROSITE; PS50275; SAC; 1.  
KW Hydrolase; Alternative splicing; Repeat; Endocytosis; RNA-binding;  
KW Multigene family.  
FT DOMAIN 119 442 SAC.  
FT DOMAIN 500 899 CATALYTIC (POTENTIAL).  
FT DOMAIN 902 971 RNA-BINDING (RRM).  
FT DOMAIN 900 1575 PRO-RICH.  
FT DOMAIN 1033 1036 POLY-SER.  
FT DOMAIN 1108 1113 POLY-PRO.  
FT DOMAIN 1126 1129 POLY-PRO.  
FT DOMAIN 1487 1490 POLY-GLU.  
FT DOMAIN 1540 1546 POLY-PRO.  
FT DOMAIN 1396 1419 3 X 3 AA REPEATS OF N-P-F.  
FT REPEAT 1396 1398 1.  
FT REPEAT 1406 1408 2.  
FT REPEAT 1417 1419 3.  
FT VARSPLIC 1306 1311 VKTNGI -> QEOPSG (in isoform 2).  
FT VARSPLIC 1312 1575 /FTid=VSP 002682.  
FT VARSPLIC 1575 Missing (in isoform 2).  
FT VARSPLIC 1575 /FTid=VSP 002683.  
SQ SEQUENCE 1575 AA; 173345 MW; 50646F6CC043B9E7 CRC64;  
  
Query Match 8.2%; Score 76; DB 1; Length 1575;  
Best Local Similarity 22.6%; Pred. No. 31;  
Matches 38; Conservative 21; Mismatches 49; Indels 60; Gaps 8;  
  
QY 8 PVLHLVALNTPLS--GGMRGIRGA-----DFQCQOAR-AVGLSGTFRAPLSSRLQ 55  
Db 648 PFIRDVAVDVTKMGCGATGNKGAVAIRMLFHTTSLCFVCSHFAAGSQ-----VKERNE 702  
  
QY 56 DLYSIVRRAD-----RGSVPIVNLKDEVLSPSWDSLFGSQ--- 91  
Db 703 DFIEIARKLSFPMGRMLFSDHYVFWCGDFNYRIDLPNEEVKELIRQQNWDSLIGDQLIN 762  
  
QY 92 ----GQLQPG-----ARIFS--FDGRDVLRRHPAWPKSVW 120  
Db 763 QKNAGQVFRGFLGKVTFAPTYKYDLFSDDDYDTSEKCRTPAWTDRLW 810  
  
RESULT 10  
CATA\_METBA STANDARD; PRT; 505 AA.  
AC O93662;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Catalase (EC 1.11.1.6).

GN KAT.  
OS Methanosarcina barkeri.  
OC Archaea; Euryarchaeota; Euryarchaeota orders incertae sedis;  
OC Methanosarcinales; Methanosarcinaceae; Methanosarcina.  
OX NCBI\_TaxID=2208;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Fusaro / DSM 804;  
RX MEDLINE=99311271; PubMed=10382262;  
RA Shima S., Netrusov A., Sordel M., Wicke M., Hartmann G.C.,  
RA Thauer R.K.;  
RT "Purification, characterization, and primary structure of a  
RT nonfunctional catalase from Methanosarcina barkeri.";  
RL Arch. Microbiol. 171:317-323(1999).  
CC -!- FUNCTION: Decomposes hydrogen peroxide into water and oxygen;  
CC serves to protect cells from the toxic effects of hydrogen  
CC peroxide.  
CC -!- CATALYTIC ACTIVITY: 2 H(2)O(2) = O(2) + 2 H(2)O.  
CC -!- COFACTOR: HEME GROUP (BY SIMILARITY).  
CC -!- SIMILARITY: Belongs to the catalase family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AJ005939; CAA06774.1; --  
DR HSSP; P04040; 1F4J.  
DR InterPro; IPR002226; Catalase.  
DR Pfam; PF00199; catalase; 1.  
DR PRINTS; PR00067; CATALASE.  
DR ProDom; PD000510; Catalase; 1.  
DR PROSITE; PS00437; CATALASE\_1; 1.  
DR PROSITE; PS00438; CATALASE\_2; 1.  
KW Oxidoreductase; Peroxidase; Iron; Heme; Hydrogen peroxide.  
FT ACT\_SITE 58 58 BY SIMILARITY.  
FT ACT\_SITE 131 131 BY SIMILARITY.  
FT METAL 341 341 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).  
SQ SEQUENCE 505 AA; 57065 MW; 2A27C4BEC47BE854 CRC64;  
  
Query Match 8.1%; Score 75; DB 1; Length 505;  
Best Local Similarity 20.7%; Pred. No. 9.8;  
Matches 30; Conservative 30; Mismatches 47; Indels 38; Gaps 7;  
  
QY 35 QQARAVGLSGTFRAPLSSRLQDLYSIVRRADRGSVPIVNLKDEVLSPSWDSLFGSQQL 94  
Db 231 EEAEKIGGSDPDHA-----TRDLYEAIK---KGDYPSWTLENQIMTPE----- 270  
  
QY 95 QPGARIFSGDGRDVLRRHPAWPKSVVHGSDDPS---GRRLMESYCETWRTTGTATGQASS 151  
Db 271 --QAEDYRFDIRDITK--VWP-----HGDFPTMKIGKLVLRNPTNYFAEVEQAAPSPAN 321  
  
QY 152 LLSG-----RLLEQKAASCHNSYI 170  
Db 322 LVPGIGISPDKMLQGRVFSYHDTI 346  
  
RESULT 11  
SYJ1\_BOVIN STANDARD; PRT; 1324 AA.  
ID SYJ1\_BOVIN  
AC O18964;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Synaptotagmin 1 (EC 3.1.3.36) (Synaptic inositol-1,4,5-trisphosphate 5-  
DE phosphatase 1) (p150) (Fragment).  
GN SYNJ1.  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;



OC Bovidae; Bovinae; Bos.  
OX NCBI\_TaxID=9913;  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 321-339 AND 454-469.  
RC TISSUE=Brain;  
RX MEDLINE=97342621; PubMed=9199318;  
RA Sakisaka T., Itoh T., Miura K., Takenawa T.;  
RT "Phosphatidylinositol 4,5-bisphosphate phosphatase regulates the  
RL rearrangement of actin filaments.";  
RL Mol. Cell. Biol. 17:3841-3849(1997).  
CC -!- FUNCTION: Hydrolyzes PIP2 bound to actin regulatory proteins  
CC resulting in the rearrangement of actin filaments downstream of  
CC tyrosine kinase and ASH/GRB2.  
CC -!- CATALYTIC ACTIVITY: 1-phosphatidyl-1D-myo-inositol 4,5-  
CC bisphosphate + H(2)O = 1-phosphatidyl-1D-myo-inositol 4-phosphate  
CC + phosphate.  
CC -!- SUBUNIT: BINDS TO AMPHIPHYSIN AND ASH/GRB2.  
CC -!- SUBCELLULAR LOCATION: PREDOMINANTLY CONCENTRATED IN THE  
CC PERINUCLEAR AREAS.  
CC -!- TISSUE SPECIFICITY: UBIQUITOUSLY EXPRESSED WITH HIGHEST LEVELS IN  
CC BRAIN.  
CC -!- DOMAIN: THE C-TERMINAL PROLINE-RICH REGION MEDIATES BINDING TO A  
CC VARIETY OF SH3 DOMAIN-CONTAINING PROTEINS INCLUDING AMPHIPHYSIN,  
CC AND ASH/GRB2.  
CC -!- SIMILARITY: IN THE CENTRAL SECTION; BELONGS TO THE INOSITOL-1,4,5-  
CC TRISPHOSPHATE 5-PHOSPHATASE FAMILY.  
CC -!- SIMILARITY: Contains 1 SAC domain.  
CC -!- SIMILARITY: Contains 1 RNA recognition motif (RRM) domain.  
CC -!- CAUTION: THIS IS A CONCEPTUAL TRANSLATIONS; A NUMBER OF POTENTIAL  
CC FRAMESHIFTS WERE CORRECTED STARTING IN POSITION 1213 SO AS TO  
CC EXTEND THE SIMILARITY WITH THE ORTHOLOGS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; D85682; BAA21652.1; ALT FRAME.  
DR InterPro; IPR005135; Exo\_endo\_phos.  
DR InterPro; IPR000300; IPPC.  
DR InterPro; IPR000504; RNA\_rec\_mot.  
DR InterPro; IPR002013; Syja\_N.  
DR Pfam; PF03372; Exo\_endo\_phos; 1.  
DR Pfam; PF02383; Syja\_N; 1.  
DR SMART; SM00128; IPPC; 1.  
DR PROSITE; PS50102; RRM; 1.  
DR PROSITE; PS50275; SAC; 1.  
DR Hydrolase; Endocytosis; RNA-binding; Multigene family.  
KW FT DOMAIN 119 442  
FT DOMAIN 475 859  
FT DOMAIN 902 971  
FT DOMAIN 860 1212  
FT DOMAIN 1033 1036  
FT DOMAIN 1108 1113  
FT DOMAIN 1126 1129  
FT CONFLICT 335 335  
FT NON\_TER 1324 1324  
SQ SEQUENCE 1324 AA; 146582 MW; EDDC2DD9D6E3191C CRC64;  
  
Query Match 8.1%; Score 75; DB 1; Length 1324;  
Best Local Similarity 22.6%; Pred. No. 31;  
Matches 38; Conservative 20; Mismatches 50; Indels 60; Gaps 8;  
  
QY 8 PVJLHVALNTPLS--GGMEGIRGA-----DFQCFOQAR-AVGLSGTFRFLSLRLQ 55  
Db 648 PFIRDVAVDVTKTGMGGATGNKGAVAIRMLFHTTSLCFVCSHFAGQSQ-----VKERN 702  
QY 56 DLYSIVRRAD-----RGSVPIVNLKDEVLSFSDLSFGSQ--- 91  
Db 703 DFLEIARKLSFPMGRLLFSDHYVFCGDFNVDLPNEEVKELIRQONWDSLIAGDQLIN 762

QY 92 -----GQLQPG-----ARIFS--FDGRDVLRLHPAWEQKSVW 120  
Db 763 QKNAGQIFRGFLEGGKVTFTPTKYDLEFSDDDYDTSEKCRTPAWTRVLW 810  
  
RESULT 12  
RIAB CVHSA  
ID RIAB CVHSA STANDARD; PRT; 7073 AA.  
AC PS9641;  
DT 15-SEP-2003 (Rel. 42, Created)  
DT 15-SEP-2003 (Rel. 42, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Replicase polyprotein lab (pp1ab) (ORF1AB) [Includes: Replicase  
DE polyprotein 1a (pp1a) (ORF1A)] [Contains: Leader protein; p65 homolog;  
DE Papain-like proteinase (EC 3.4.24.-) (NSP1); 3C-like proteinase  
DE (EC 3.4.24.-) (3CL-PRO) (NSP2); HD2 (NSP3); NSP4; NSP5; Growth  
DE factor-like (NSP7); RNA-directed RNA polymerase (EC 2.7.7.48) (RdRp)  
DE (NSP9); Helicase (Hel) (NSP10); NSP11; NSP12; NSP13].  
OS Human coronavirus (strain SARS) (HCoV-SARS).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;  
OC Coronaviridae; Coronavirus.  
OX NCBI\_TaxID=227859;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Isolate Urbani;  
RA Bellini W.J., Campagnoli R.P., Icenogle J.P., Monroe S.S., Nix W.A.,  
RA Oberste M.S., Pallansch M.A., Rota P.A.;  
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Isolate Tor2;  
RA Marra M., Jones S.J.M., Holt R.;  
RT "The complete genome of the SARS associated coronavirus.";  
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Isolate CUHK-W1;  
RA Tsui S.K.W., Lo D.Y.M., Tam J.S., Fung K.P., Chim S.S.C., Au C.C.,  
RA Chan A.H., Wan A.W.K., Au K.W., Chan C.W., Kou C.Y.C., Lam H.M.,  
RA Lam W.Y., Lau S.K., Lau Y.L., Lau Y.M., Law S.L., Law T.W., Li M.L.Y.,  
RA Tse C.H., Wong C.H., Yiu W.H., Lee C.Y., Chan A.K.C., Chiu R.W.K.,  
RA Ng E.K.O., Tong Y.K., Chan P.K.S., Au-Yeung C., Cheung J.K.L., Chu I.,  
RA Hung E.C.W., Waye M.M.Y.;  
RT "DNA sequence of a human coronavirus (CUHK-W1) from a patient with  
RT severe acute respiratory syndrome (SARS) in Hong Kong.";  
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Isolate HKU-39849;  
RA Leung F.C., Zeng F., Chan C.W.M., Chan C.M.Y., Chen J., Chow K.Y.C.,  
RA Hon C.C.C., Hui R.K.H., Li J., Li V.Y.Y., Wang Y.Y., Peiris J.S.M.,  
RA Poon L.L.M.;  
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RP SEQUENCE OF 4993-5127 FROM N.A.  
RC STRAIN=Isolate Vietnam;  
RA Emery S., Erdman D., Peret T., Ksiazek T.;  
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.  
RN [6]  
RP SEQUENCE OF 4993-5136 FROM N.A.  
RC STRAIN=Isolate Taiwan;  
RA Lin J.-H., Chiu S.-C., Yang J.-Y., Wang S.-F., Chen H.-Y.;  
RT "Detection of a novel human coronavirus in a severe acute respiratory  
RT syndrome patient in Taiwan.";  
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: The replicase polyprotein of coronaviruses is a  
CC multifunctional protein: it contains the activities necessary for  
CC the transcription of negative stranded RNA, leader RNA, subgenomic  
CC mRNAs and progeny virion RNA as well as proteinases responsible  
CC for the cleavage of the polyprotein into functional products (By  
CC similarity).  
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +





```
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; UI7227; AAA90984.1; ALT_INIT.
DR EMBL; AF031938; AAC61899.1; -.
DR EMBL; AL591788; CAC46304.1; -.
KW PHB biosynthesis; Transferase; Acyltransferase; Complete proteome.
FT ACT_SITE 349 349 POTENTIAL.
FT CONFLICT 111 111 G -> A (IN REF. 1).
FT CONFLICT 122 122 R -> H (IN REF. 1).
FT CONFLICT 157 157 R -> K (IN REF. 1).
FT CONFLICT 477 477 K -> R (IN REF. 1).
FT CONFLICT 481 481 V -> M (IN REF. 1).
SQ SEQUENCE 611 AA; 68039 MW; ABE8A9B17F87D7A9 CRC64;

Query Match 8.1%; Score 74.5; DB 1; Length 611;
Best Local Similarity 27.4%; Pred. No. 14;
Matches 40; Conservative 16; Mismatches 47; Indels 43; Gaps 9;

QY 49 FLSSRLQ--DLYSIVRRADRGSV--PIVNL--KDEVLSPSWDSLFSGSQGLQPGARIFS 102
Db 470 YLENRLSKGEMVLAGRVSGLGVKIPYLNATKEDHIAPA-KSVFLGSS-----S 518

QY 103 FDGR-----DVLRRHPAPQKSVVHGHSDPSGRRLMESYCETW--RTTETGA-- 145
Db 519 FGGKVTFLVSGSGHIAGVNPPARSKYQYWTGGAPKGD-----IETWMGKAKETAGSWM 572

QY 146 ---TGQASSLLSGRLLEQKAASCHNS 168
Db 573 PHWQGWVERLDKRRVPARKAGGPLNS 598

RESULT 14
P47K_PSECL STANDARD; PRT; 419 AA.
AC P31521;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 47 kDa protein (P47K).
OS Pseudomonas chlororaphis (Pseudomonas aureofaciens).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=333;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B23;
RX MEDLINE=91193202; PubMed=2013568;
RA Nishiyama M., Horinouchi S., Kobayashi M., Nagasawa T., Yamada H.,
RA Beppu T.
RT "Cloning and characterization of genes responsible for metabolism of
RT nitrile compounds from Pseudomonas chlororaphis B23.";
RL J. Bacteriol. 173:2465-2472(1991).
CC -!- FUNCTION: THE PRESENCE OF P47K IS CRITICAL FOR THE EXPRESSION
CC OF THE NITRILE HYDRATASE GENES. MAY STABILIZE OR ACTIVATE THE
CC NITRILE HYDRATASE PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D90216; BAA14247.1; -.
DR PIR; D42725; D42725.
DR InterPro; IPR003495; CobW.
DR Pfam; PF02492; cobW; 1.
SQ SEQUENCE 419 AA; 46666 MW; FF5113800E27FF0C CRC64;

Query Match 7.9%; Score 72.5; DB 1; Length 419;
Best Local Similarity 22.1%; Pred. No. 14;
Matches 40; Conservative 24; Mismatches 60; Indels 57; Gaps 8;
```

```
QY 26 IRGADFQCFQQAARAVGLSGTFRFLSSR-LQDLYSIVRRADRGSVPIVNLKDEVLSPSWD 84
Db 141 VDGSGFQALLESTDTVARADTEAHTSTRHLADL--LIEQVEYANVILVKNRDLIDEPGYQ 198

QY 85 SLFSGSQGLQPGARIFS-----FDGRDVLRRHPAPQK-----SVWHGS 123
Db 199 AVHAILAG-LNPSARIMPMAGHGNVALSSLLDTHLFDLPSLAASPQWMRKMEATDTPASES 257

QY 124 DPSG-----RRLMESYCETWRTTETTTGATGQASSLLSGRLLEQKA-----ASC 165
Db 258 DTGYVTSWVYRERAPFPHPQRLLEFLQKPWH-----NGRLLRSKGYFWLASR 303

QY 166 H 166
Db 304 H 304

RESULT 15
VNUC_EBOG4 STANDARD; PRT; 739 AA.
AC Q9QCE9;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Nucleoprotein (Nucleocapsid protein).
GN NP.
OS Ebola virus (strain Gabon-94) (Ebo).
OC Viruses; ssRNA negative-strand viruses; Mononegavirales; Filoviridae;
OC Ebola-like viruses.
OX NCBI_TaxID=128947;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99036017; PubMed=9820131;
RA Prehaud C.J.C., Hellebrand E., Coudrier D., Volchkov V.E.,
RA Volchkova V.A., Feldmann B., Le Guenne B., Bouloy M.;
RT "Recombinant Ebola virus nucleoprotein and glycoprotein (Gabon 94
RT strain) provide new tools for the detection of human infections.";
RL J. Gen. Virol. 79:2565-2572(1998).
CC -!- FUNCTION: RESPONSIBLE FOR ENCAPSIDATION OF GENOMIC RNA.
CC -!- DOMAIN: THIS PROTEIN CAN BE DIVIDED INTO A HYDROPHOBIC N-TERMINAL
CC HALF, AND A HYDROPHILIC AND HIGHLY ACIDIC C-TERMINAL HALF.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Y09358; CAA70541.1; -.
KW Nucleoprotein.
SQ SEQUENCE 739 AA; 83240 MW; 33A96720FFBC638E CRC64;

Query Match 7.9%; Score 72.5; DB 1; Length 739;
Best Local Similarity 27.0%; Pred. No. 27;
Matches 43; Conservative 16; Mismatches 67; Indels 33; Gaps 8;

QY 22 GMRGIRGAD-----FQCFQQAARAVGLSGTFRFLSSRLQD-----LYSIVRRADRGSV 69
Db 218 GMMVAGHDANDAVISNSVAQAFSGLL-IVKTVLDHILQKTQGVRLHPLARTA----- 271

QY 70 PIVNLKDEV--LSPSWDSLFSGSQGLQPGARIFSFDGRDVLRRHPAPQKS-----VW 120
Db 272 ---KVRNEVNSLKAALSSL--AKHGEYAPPARLLNLSGVNLEHGLFPQLSAIALGVATA 326

QY 121 HGSDPSGRRLMESYCETWRTTETTTGATGQASSLLSGRLLE 159
Db 327 HGSTLAGNVGEQY-QQLREAATEAEKQLQQYAESRELD 364

Search completed: February 17, 2004, 09:57:04
```

Job time : 19 secs

---



Db 121 HGSDPSGRRRLMESYCETWRTTGTGATGQASSLLSGRLLLEQKAASCHNSYIVLCIE 175

RESULT 2

Q61434 PRELIMINARY; PRT; 1140 AA.

AC Q61434; Q60672;

DT 01-NOV-1996 (TReMBLrel. 01, Created)

DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)

DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)

DE Collagen (Fragment).

GN COL15A1.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RA Abe N., Muragaki Y., Yoshioka H., Inoue H., Ninomiya Y.;

RT "Identification of a novel collagen chain represented by extensive

RT interruptions in the triple-helical region.";

RL Cell. Mol. Biol. Res. 196:576-582(1993).

DR EMBL; D17546; BAA04483.1; -.

DR HSSP; P39061; IKOE.

DR MGD; MGI:88449; Coll15a1.

DR InterPro; IPR000087; Collagen.

DR InterPro; IPR003129; TSPN.

DR Pfam; PF01391; Collagen; 8.

DR Pfam; PF02210; TSPN; 1.

DR ProDom; PD000007; Collagen; 1.

KW Collagen.

FT NON\_TER

SQ SEQUENCE 1140 AA; 115156 MW; 880C7E6862B3BD FE CRC64;

Query Match 100.0%; Score 923; DB 11; Length 1140;

Best Local Similarity 100.0%; Pred. No. 2.8e-83;

Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HTHQDFQPVLHLVALNTPLSGMGRGADFCFQQARAVGLSGTFRFLSSRLQDLYSI 60

Db 957 HTHQDFQPVLHLVALNTPLSGMGRGADFCFQQARAVGLSGTFRFLSSRLQDLYSI 1016

QY 61 VRRADRGSVPIVNLKDEVLSFSGSQQLPGARIFSDGRDVLRHHPWPQKSVW 120

Db 1017 VRRADRGSVPIVNLKDEVLSFSGSQQLPGARIFSDGRDVLRHHPWPQKSVW 1076

QY 121 HGSDPSGRRRLMESYCETWRTTGTGATGQASSLLSGRLLLEQKAASCHNSYIVLCIE 175

Db 1077 HGSDPSGRRRLMESYCETWRTTGTGATGQASSLLSGRLLLEQKAASCHNSYIVLCIE 1131

RESULT 3

Q62001 PRELIMINARY; PRT; 1774 AA.

AC Q62001; Q60672;

DT 01-NOV-1996 (TReMBLrel. 01, Created)

DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)

DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)

DE Procollagen, type XVIII, alpha 1 precursor (XVIII) collagen

DE (Procollagen, type XVIII, alpha 1) (Alpha-1 type XVIII collagen).

GN COL18A1.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=PRIMARY TAIL CULTURE;

RX MEDLINE=94245707; PubMed=8188673;

RA Rehn M., Hintikka E., Pihlajaniemi T.;

RT "Primary structure of the alpha 1 chain of mouse type XVIII collagen,

RT partial structure of the corresponding gene, and comparison of the

RT alpha 1(XVIII) chain with its homologue, the alpha 1(XV) collagen

chain.";

RL J. Biol. Chem. 269:13929-13935(1994).

RN [2]

RP SEQUENCE OF 1-562 FROM N.A.

RX MEDLINE=94240112; PubMed=8183894;

RA Rehn M., Pihlajaniemi T.;

RT "Alpha 1(XVIII), a collagen chain with frequent interruptions in the

RT collagenous sequence, a distinct tissue distribution, and homology

RT with type XV collagen.";

RL Proc. Natl. Acad. Sci. U.S.A. 91:4234-4238(1994).

RN [3]

RP SEQUENCE OF 1-562 FROM N.A.

RX MEDLINE=95181468; PubMed=7876242;

RA Rehn M., Pihlajaniemi T.;

RT "Identification of three N-terminal ends of type XVIII collagen chains

RT and tissue-specific differences in the expression of the corresponding

RT transcripts. The longest form contains a novel motif homologous to rat

RT and Drosophila frizzled proteins.";

RL J. Biol. Chem. 270:4705-4711(1995).

DR EMBL; U03715; AAC52903.1; -.

DR EMBL; U03716; AAC52903.1; JOINED.

DR EMBL; U03718; AAC52903.1; JOINED.

DR EMBL; U34607; AAC52903.1; JOINED.

DR EMBL; U34608; AAC52903.1; JOINED.

DR EMBL; U34609; AAC52903.1; JOINED.

DR EMBL; U34610; AAC52903.1; JOINED.

DR EMBL; U34611; AAC52903.1; JOINED.

DR EMBL; U34612; AAC52903.1; JOINED.

DR EMBL; U34613; AAC52903.1; JOINED.

DR EMBL; U11637; AAC52179.1; -.

DR HSSP; P39061; IKOE.

DR MGD; MGI:88451; Coll18a1.

DR InterPro; IPR000087; Collagen.

DR InterPro; IPR000024; Fz\_domain.

DR InterPro; IPR001791; Laminin\_G.

DR InterPro; IPR003129; TSPN.

DR Pfam; PF01391; Collagen; 8.

DR Pfam; PF01392; Fz; 1.

DR Pfam; PF02210; TSPN; 1.

DR ProDom; PD000007; Collagen; 1.

DR SMART; SM00063; FRI; 1.

DR SMART; SM00282; LamG; 1.

DR SMART; SM00210; TSPN; 1.

DR PROSITE; PS50038; FZ; 1.

KW Collagen; Signal.

SQ SEQUENCE 1774 AA; 182229 MW; CF4D9BC9E88EF232 CRC64;

Query Match 100.0%; Score 923; DB 11; Length 1774;

Best Local Similarity 100.0%; Pred. No. 5e-83;

Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HTHQDFQPVLHLVALNTPLSGMGRGADFCFQQARAVGLSGTFRFLSSRLQDLYSI 60

Db 1591 HTHQDFQPVLHLVALNTPLSGMGRGADFCFQQARAVGLSGTFRFLSSRLQDLYSI 1650

QY 61 VRRADRGSVPIVNLKDEVLSFSGSQQLPGARIFSDGRDVLRHHPWPQKSVW 120

Db 1651 VRRADRGSVPIVNLKDEVLSFSGSQQLPGARIFSDGRDVLRHHPWPQKSVW 1710

QY 121 HGSDPSGRRRLMESYCETWRTTGTGATGQASSLLSGRLLLEQKAASCHNSYIVLCIE 175

Db 1711 HGSDPSGRRRLMESYCETWRTTGTGATGQASSLLSGRLLLEQKAASCHNSYIVLCIE 1765

RESULT 4

Q9QZD2

ID Q9QZD2 PRELIMINARY; PRT; 226 AA.

AC Q9QZD2;

DT 01-MAY-2000 (TReMBLrel. 13, Created)

DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)

DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)

DE Collagen XVIII (Fragment).

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Sprague-Dawley;  
RX MEDLINE=20227226; PubMed=10766159;  
RA Perletti G., Concarì P., Giardini R., Marras E., Piccinini F.,  
RA Folkman J., Chen L.;  
RT "Antitumor activity of endostatin against carcinogen-induced rat  
RT primary mammary tumors.";  
RL Cancer Res. 60:1793-1796(2000).  
DR EMBL; AF189709; AAF00975.1; -.  
DR HSSP; P39061; IKOE.  
FT NON\_TER 1  
SQ SEQUENCE 226 AA; 25350 MW; 38B83C0486C0E949 CRC64;  
  
Query Match 95.6%; Score 882; DB 11; Length 226;  
Best Local Similarity 95.4%; Pred. No. 4.5e-80;  
Matches 167; Conservative 1; Mismatches 7; Indels 0; Gaps 0;  
  
QY 1 HTHQDFQPVHLVALNTPLSGMVGIRGADFCFQOQARAVGLSGTFRFLSSRLQDLYSI 60  
Db 43 HTHQDFHPVHLVALNTPLSGMVGIRGADFCFQOQARAVGLSGTFRFLSSRLQDLYSI 102  
  
QY 61 VRRADRGSPVIVNLKDEVLSWDSLFSGSQGLQPGARIFSFDRDVLRLHPAPQKSVW 120  
Db 103 VRRADRGSPVIVNLKDEVLSWDSLFSGSQGLHSGARIFSFDRDVLRLHPAPQKSVW 162  
  
QY 121 HGSDPSGRRLMESYCETWRTTGTATGQASSLLSGRLLEQKAASCHNSYIVLCIE 175  
Db 163 HGSDPSGRRLMESYCETWRTTGTATGQASSLLSGRLLEQKAASCHNSYIVLCIE 217

RESULT 5  
Q9WUW5  
ID Q9WUW5 PRELIMINARY; PRT; 171 AA.  
AC Q9WUW5;  
DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE Collagen type XVIII, alpha (I) chain (Fragment).  
GN COL18A1.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Liver;  
RA Jia J.D., Bauer M., Eberspaecher U., Donner P., Schuppan D.;  
RT "Partial 3'-terminal cDNA sequence of rat collagen XVIII/endostatin.";  
RT Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Liver;  
RA Jia J.D., Bauer M., Sedlacek N., Ruehl M., Riecken E.O., Schuppan D.;  
RT "Temporospatial expression of collagen XVIII/endostatin in acute and  
RT chronic liver injuries.";  
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ236873; CAB44263.1; -.  
DR HSSP; P39061; IKOE.  
FT NON\_TER 1  
FT NON\_TER 171  
SQ SEQUENCE 171 AA; 18933 MW; 81BE2EE3FC2C8E72 CRC64;  
  
Query Match 90.7%; Score 837; DB 11; Length 171;  
Best Local Similarity 95.8%; Pred. No. 9.8e-76;  
Matches 160; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 9 VHLVALNTPLSGMVGIRGADFCFQOQARAVGLSGTFRFLSSRLQDLYSIVRRADRS 68  
Db 1 VHLVALNTPLSGMVGIRGADFCFQOQARAVGLSGTFRFLSSRLQDLYSIVRRADRS 60

QY 69 VPIVNLKDEVLSWDSLFSGSQGLQPGARIFSFDRDVLRLHPAPQKSVHGSDDPSGR 128  
Db 61 VPIVNLKDEVLSWDSLFSGSQGLHSGARIFSFDRDVLRLHPAPQKSVHGSDDPSGR 120  
  
QY 129 RLMESYCETWRTTGTATGQASSLLSGRLLEQKAASCHNSYIVLCIE 175  
Db 121 RLMESYCETWRTTGTATGQASSLLSGRLLEQKAASCHNSYIVLCIE 167  
  
RESULT 6  
Q8WXI5  
ID Q8WXI5 PRELIMINARY; PRT; 187 AA.  
AC Q8WXI5;  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
DE Collagen XVIII (Fragment).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21409408; PubMed=11517600;  
RA Feng Y., Cui L.B., Liu C.X., Ma Q.J.;  
RT "Inhibition effect in vitro of purified endostatin expressed in Pichia  
RT pastoris.";  
RL Sheng Wu Gong Cheng Xue Bao 17:278-282(2001).  
DR EMBL; AF416592; AAL37720.1; -.  
FT NON\_TER 1  
SQ SEQUENCE 187 AA; 20448 MW; 72B1047D85838CD3 CRC64;  
  
Query Match 88.1%; Score 813; DB 4; Length 187;  
Best Local Similarity 85.7%; Pred. No. 2.8e-73;  
Matches 150; Conservative 14; Mismatches 11; Indels 0; Gaps 0;

QY 1 HTHQDFQPVHLVALNTPLSGMVGIRGADFCFQOQARAVGLSGTFRFLSSRLQDLYSI 60  
Db 5 HSHRDFQPVHLVALNSPLSGMVGIRGADFCFQOQARAVGLAGTFRFLSSRLQDLYSI 64  
  
QY 61 VRRADRGSPVIVNLKDEVLSWDSLFSGSQGLQPGARIFSFDRDVLRLHPAPQKSVW 120  
Db 65 VRRADRGSPVIVNLKDEVLSWDSLFSGSQGLHSGARIFSFDRDVLRLHPAPQKSVW 124  
  
QY 121 HGSDPSGRRLMESYCETWRTTGTATGQASSLLSGRLLEQKAASCHNSYIVLCIE 175  
Db 125 HGSDPNGRRLTESYCETWRTTGTATGQASSLLSGRLLEQKAASCHNSYIVLCIE 179

RESULT 7  
Q8NGI9  
ID Q8NGI9 PRELIMINARY; PRT; 261 AA.  
AC Q8NGI9;  
DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Multi-functional protein MFP.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Dou D.;  
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF282883; AAM52249.1; -.  
SQ SEQUENCE 261 AA; 26745 MW; CA60C920AF3E90E5 CRC64;  
  
Query Match 88.1%; Score 813; DB 4; Length 261;  
Best Local Similarity 85.7%; Pred. No. 4.2e-73;  
Matches 150; Conservative 14; Mismatches 11; Indels 0; Gaps 0;



QY 1 HTHQDFQPVHLVNTPLSGMIRGADFCQFQQAARAVGLSGTFRFLSSRLQDLYSI 60  
Db 79 HSHRDFQPVHLVNTPLSGMIRGADFCQFQQAARAVGLSGTFRFLSSRLQDLYSI 138  
QY 61 VRRADRGSVPIVNLKDEVLSWDSFSGSQQLQPGARIPFSDGRDVLHRHPWPQKSVW 120  
Db 139 VRRADRAAVPIVNLKDELLFPFSEALFSGSEGLPKPGARIPFSDGKDVLRHPTWPQKSVW 198  
QY 121 HGSDPSGRRRLMESYCETWRTTGTGATGQASSLLSGRLLEQKAASCHNSYIVLCIE 175  
Db 199 HGSDPNGRRLTESYCETWRTTGTGATGQASSLLSGRLLEQKAASCHNSYIVLCIE 253  
RESULT 8  
Q8N4S4 PRELIMINARY; PRT; 816 AA.  
AC Q8N4S4;  
DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Similar to collagen, type XVIII, alpha 1 (Fragment).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Renal adenocarcinoma;  
RA Strausberg R.;  
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC033715; AAH33715.1; -.  
DR InterPro; IPR000087; Collagen.  
DR Pfam; PF01391; Collagen; 6.  
DR ProDom; PD000007; Collagen; 1.  
KW Collagen.  
FT NON TER .1 1  
SQ SEQUENCE 816 AA; 82553 MW; 5D539B2946694F86 CRC64;  
Query Match 88.1%; Score 813; DB 4; Length 816;  
Best Local Similarity 85.7%; Pred. No. 1.8e-72;  
Matches 150; Conservative 14; Mismatches 11; Indels 0; Gaps 0;  
QY 1 HTHQDFQPVHLVNTPLSGMIRGADFCQFQQAARAVGLSGTFRFLSSRLQDLYSI 60  
Db 634 HSHRDFQPVHLVNTPLSGMIRGADFCQFQQAARAVGLSGTFRFLSSRLQDLYSI 693  
QY 61 VRRADRGSVPIVNLKDEVLSWDSFSGSQQLQPGARIPFSDGRDVLHRHPWPQKSVW 120  
Db 694 VRRADRAAVPIVNLKDELLFPFSEALFSGSEGLPKPGARIPFSDGKDVLRHPTWPQKSVW 753  
QY 121 HGSDPSGRRRLMESYCETWRTTGTGATGQASSLLSGRLLEQKAASCHNSYIVLCIE 175  
Db 754 HGSDPNGRRLTESYCETWRTTGTGATGQASSLLSGRLLEQKAASCHNSYIVLCIE 808  
RESULT 9  
Q9CRT2 PRELIMINARY; PRT; 160 AA.  
AC Q9CRT2;  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE Procollagen, type XVIII, alpha 1 (Fragment).  
GN COL18A1.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Embryo;  
RX MEDLINE=21085660; PubMed=11217851;  
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,

RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,  
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,  
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,  
RA Hayashizaki Y.;  
RT "Functional annotation of a full-length mouse cDNA collection."  
RL Nature 409:685-690(2001).  
DR EMBL; AK014292; BAB29249.1; -.  
DR HSSP; P39061; 1KOE.  
DR MGD; MGI:88451; Coll18a1.  
FT NON TER 1  
SQ SEQUENCE 160 AA; 17725 MW; 60F853D777C375D2 CRC64;  
Query Match 85.9%; Score 793; DB 11; Length 160;  
Best Local Similarity 100.0%; Pred. No. 2.2e-71;  
Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 25 GIRGADFCQFQQAARAVGLSGTFRFLSSRLQDLYSIVRRDRGSVPIVNLKDEVLSWSD 84  
Db 1 GIRGADFCQFQQAARAVGLSGTFRFLSSRLQDLYSIVRRDRGSVPIVNLKDEVLSWSD 60  
QY 85 SLFSGSQQLQPGARIPFSDGRDVLHRHPWPQKSVWHGSDPSGRRRLMESYCETWRTTGTG 144  
Db 61 SLFSGSQQLQPGARIPFSDGRDVLHRHPWPQKSVWHGSDPSGRRRLMESYCETWRTTGTG 120  
QY 145 ATGQASSLLSGRLLEQKAASCHNSYIVLCIE 175  
Db 121 ATGQASSLLSGRLLEQKAASCHNSYIVLCIE 151  
RESULT 10  
O93419 PRELIMINARY; PRT; 1344 AA.  
AC O93419;  
DT 01-NOV-1998 (TrEMBLrel. 08, Created)  
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Collagen XVIII precursor.  
OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
OC Gallus.  
OX NCBI\_TaxID=9031;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98411346; PubMed=9738008;  
RA Halfter W., Dong S., Schurer B., Cole G.J.;  
RT "Collagen XVIII is a basement membrane heparan sulfate proteoglycan."  
RL J. Biol. Chem. 273:25404-25412(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Halfter W., Dong S.;  
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF083440; AAC33294.2; -.  
DR HSSP; P39061; 1KOE.  
DR InterPro; IPR000087; Collagen.  
DR InterPro; IPR001791; Laminin\_G.  
DR InterPro; IPR003129; TSPN.  
DR Pfam; PF01391; Collagen; 8.  
DR Pfam; PF02210; TSPN; 1.  
DR ProDom; PD000007; Collagen; 2.

DR SMART; SM00282; LamG; 1.  
DR SMART; SM00210; TSPN; 1.  
KW Collagen; Signal.  
FT SIGNAL 1 26 POTENTIAL.  
SQ SEQUENCE 1344 AA; 137402 MW; 7AA366E4FE940CCD CRC64;  
  
Query Match 78.4%; Score 724; DB 13; Length 1344;  
Best Local Similarity 76.0%; Pred. No. 2.7e-63;  
Matches 133; Conservative 20; Mismatches 22; Indels 0; Gaps 0;  
  
QY 1 HTHQDFQPVHLVALNTPLSGMGRGIRGADFCFQQAQAVGLSGTFRFLSSRLQDLYSI 60  
DB 1161 HVHQDFQPALHLVALNTPLSGMGRGIRGADFCFQQAQAVGLSGTFRFLSSRLQDLYSI 1220  
  
QY 61 VRRADRGSPVIVNLKDEVLSFSGQQLPGARIFSFDRDVLRHHPWPQKSVW 120  
DB 1221 VRRADRTAVPVLNLRDEVLSFSGQQLPGARILSFDRDILQDSAMPQKSIW 1280  
  
QY 121 HGS DPGSRRLMESYCETWRTTGTATGQASSLLSGRLLEQKAASCHNSYIVLCIE 175  
DB 1281 HGSDAKGRRLPESYCEAWRTDERGTSGQASSLLSGRLLEQKAASCHNSYIVLCIE 1335  
  
RESULT 11  
Q8JFF7 PRELIMINARY; PRT; 1307 AA.  
AC Q8JFF7  
DT 01-OCT-2002 (Tremblrel. 22, Created)  
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)  
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
DE Type XVIII collagen short variant.  
OS Xenopus laevis (African clawed frog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;  
OC Xenopodinae; Xenopus.  
OX NCBI\_TaxID=8355;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=22166979; PubMed=12175494;  
RA Elam H., Peterson J., Pihlajaniemi T., Destree O.;  
RT "Cloning of three variants of type XVIII collagen and their expression  
RT patterns during Xenopus laevis development.";  
RL Mech. Dev. 114:109-113 (2002).  
DR EMBL; AY052763; AAL14257.1; -.  
DR InterPro; IPR000087; Collagen.  
DR InterPro; IPR003129; TSPN.  
DR Pfam; PF01391; Collagen; 6.  
DR Pfam; PF02210; TSPN; 1.  
DR SMART; SM00210; TSPN; 1.  
KW Collagen.  
SQ SEQUENCE 1307 AA; 134153 MW; D53EDBFE3DE34976 CRC64;  
  
Query Match 75.0%; Score 692; DB 13; Length 1307;  
Best Local Similarity 73.1%; Pred. No. 4.1e-60;  
Matches 128; Conservative 16; Mismatches 31; Indels 0; Gaps 0;  
  
QY 1 HTHQDFQPVHLVALNTPLSGMGRGIRGADFCFQQAQAVGLSGTFRFLSSRLQDLYSI 60  
DB 1124 HTHQDFNPALHLVALNPISGSMKIRGVDFQCFEQARKAGLHGTFRFLSSRLQDLYSI 1183  
  
QY 61 VRRADRGSPVIVNLKDEVLSFSGQQLPGARIFSFDRDVLRHHPWPQKSVW 120  
DB 1184 VRRADRGSPVIVNLKDEVLSFSGQQLPGARILSFDRDVLRHHPWPQKSVW 1243  
  
QY 121 HGS DPGSRRLMESYCETWRTTGTATGQASSLLSGRLLEQKAASCHNSYIVLCIE 175  
DB 1244 HGSDAKGRRLTESYCETWRTTGTATGQASSLLSGRLLEQKAASCHNSYIVLCIE 1298  
  
RESULT 12  
Q8QHL9 PRELIMINARY; PRT; 1315 AA.  
ID Q8QHL9  
AC Q8QHL9

DT 01-JUN-2002 (Tremblrel. 21, Created)  
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)  
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
DE Type XVIII collagen alpha1 chain.  
OS Xenopus laevis (African clawed frog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;  
OC Xenopodinae; Xenopus.  
OX NCBI\_TaxID=8355;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Ishino T., Sekimizu K., Natori S., Kubo T.;  
RT "Identification and characterization of genes expressed selectively in  
RT the regenerating tail of Xenopus laevis tadpole.";  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB047066; BAB84674.1; -.  
DR InterPro; IPR000087; Collagen.  
DR InterPro; IPR003129; TSPN.  
DR Pfam; PF01391; Collagen; 7.  
DR Pfam; PF02210; TSPN; 1.  
DR ProDom; PD000007; Collagen; 1.  
DR SMART; SM00210; TSPN; 1.  
KW Collagen.  
SQ SEQUENCE 1315 AA; 134946 MW; 0C56C235DE058365 CRC64;  
  
Query Match 74.8%; Score 690; DB 13; Length 1315;  
Best Local Similarity 73.1%; Pred. No. 6.6e-60;  
Matches 128; Conservative 16; Mismatches 31; Indels 0; Gaps 0;  
  
QY 1 HTHQDFQPVHLVALNTPLSGMGRGIRGADFCFQQAQAVGLSGTFRFLSSRLQDLYSI 60  
DB 1132 HTHQDFNPALHLVALNPISGSMKIRGVDFQCFEQARKAGLHGTFRFLSSRLQDLYSI 1191  
  
QY 61 VRRADRGSPVIVNLKDEVLSFSGQQLPGARIFSFDRDVLRHHPWPQKSVW 120  
DB 1192 VRRADRGSPVIVNLKDEVLSFSGQQLPGARILSFDRDVLRHHPWPQKSVW 1251  
  
QY 121 HGS DPGSRRLMESYCETWRTTGTATGQASSLLSGRLLEQKAASCHNSYIVLCIE 175  
DB 1252 HGSDAKGRRLTESYCETWRTTGTATGQASSLLSGRLLEQKAASCHNSYIVLCIE 1306  
  
RESULT 13  
Q8AWC6 PRELIMINARY; PRT; 361 AA.  
ID Q8AWC6  
AC Q8AWC6;  
DT 01-MAR-2003 (Tremblrel. 23, Created)  
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)  
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
DE Collagen XVIII (Fragment).  
GN COL18A1.  
OS Brachydanio rerio (Zebrafish) (Danio rerio).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;  
OC Cyprinidae; Danio.  
OX NCBI\_TaxID=7955;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Haftek Z., Morvan-Dubois G., Thisse B., Garrone R., Le Guellec D.;  
RT "Sequence and embryonic expression of collagen XVIII NC11 domain  
RT (endostatin) in the zebrafish.";  
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ494837; CAD38825.1; -.  
FT NON TER 1  
SQ SEQUENCE 361 AA; 40222 MW; 3C5A0F8479D26735 CRC64;  
  
Query Match 65.6%; Score 605.5; DB 13; Length 361;  
Best Local Similarity 65.1%; Pred. No. 3.5e-52;  
Matches 114; Conservative 24; Mismatches 36; Indels 1; Gaps 1;  
  
QY 1 HTHQDFQPVHLVALNTPLSGMGRGIRGADFCFQQAQAVGLSGTFRFLSSRLQDLYSI 60  
DB 179 HTHS-GPGLHLIALNSPQVGNMRGIRGADFCFQQAQAVGLSGTFRFLSSRLQDLYSI 237

Qy 61 VRRADRGSVPIVNLKDEVLSPSWDSLFRSGSQGLQPGARIEFSFGDRDVLHRHPAQKSVW 120  
| | | | | : | | | | | : | | | | | : : : | | | | | | | | | | : | | : |  
Db 238 VRRSDRETLPVNLKDQVLFSSWESLFSDESERMKDNAPIYSFGDRDVLDRDSAWPEKMIW 297  
| | | | | : | | | | | : | | | | | : : : | | | | | | | | | | : | | : |  
Qy 121 HGSDPSGRRRLMESYCETWRTTGTATGQASSLLSGRLLLEQKAASCHNSYIVLCIE 175  
| | | | | : | | | | | : | | | | | : : : | | | | | | | | | | : | | : |  
Db 298 HGSDGRGHRQTDNYCETWAGDRAVTVGLASSLQAGQLLQQTSSSCSSYIALCIE 352  
| | | | | : | | | | | : | | | | | : : : | | | | | | | | | | : | | : |  
RESULT 14  
O35206 PRELIMINARY; PRT; 1367 AA.  
AC O35206;  
DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Type XV collagen.  
GN COL15A1.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97480713; PubMed=9339358;  
RA Hagg P.M., Horelli-Kuitunen N., Eklund L., Palotie A.,  
RA Pihlajaniemi T.;  
RT "Cloning of mouse type XV collagen sequences and mapping of the  
RT corresponding gene to 4B1-3. Comparison of mouse and human alpha 1  
RT (XV) collagen sequences indicates divergence in the number of small  
RT collagenous domains";  
RL Genomics 45:31-41(1997).  
DR EMBL; AF011450; AAC53367.1; -.  
DR HSSP; P39061; 1KOE.  
DR MGD; MGI:88449; Coll15a1.  
DR InterPro; IPR000087; Collagen.  
DR InterPro; IPR001791; Laminin\_G.  
DR InterPro; IPR003129; TSPN.  
DR Pfam; PF01391; Collagen; 5.  
DR Pfam; PF02210; TSPN; 1.  
DR ProDom; PD000007; Collagen; 1.  
DR SMART; SM00282; LamG; 1.  
DR SMART; SM00210; TSPN; 1.  
KW Collagen.  
SQ SEQUENCE 1367 AA; 140525 MW; A483A1254AF3AEEC CRC64;  
Query Match 59.2%; Score 546; DB 11; Length 1367;  
Best Local Similarity 62.1%; Pred. No. 1.7e-45;  
Matches 105; Conservative 22; Mismatches 38; Indels 4; Gaps 1;  
Qy 7 QPVLHLVALNTPLSGGMRGIRGADFCQFQQAARAVGLSGTFRAPLSSRLQDLKSIYRRADR 66  
: | | | | | | | | | | : | | | | | | | | | | | | | | | | | | | : | | : |  
Db 1194 RPVLHLVALNTPVAGDIR----ADFQCFQQAARAAGLLSTFRAPLSSHLQDLSTVVRKAER 1249  
: | | | | | | | | | | : | | | | | | | | | | | | | | | | | | | : | | : |  
Qy 67 GSVPIVNLKDEVLSPSWDSLFRSGSQGLQPGARIEFSFGDRDVLHRHPAQKSVWHGSDPS 126  
: | | | | | : | | | | | : | | | | | : : : | | | | | | | | | | : | | : |  
Db 1250 FGLPIVNLKGQVLFNNWDSIFSGDGGQFNTHPIYSFGDRDVTDPSPQKVVVHGSNPH 1309  
: | | | | | : | | | | | : | | | | | : : : | | | | | | | | | | : | | : |  
RESULT 15  
Q9EQD9 PRELIMINARY; PRT; 1367 AA.  
ID Q9EQD9  
AC Q9EQD9;  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Type XV collagen.  
GN COL15A1.

OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=129/Sv;  
RX MEDLINE=97480713; PubMed=9339358;  
RA Hagg P.M., Horelli-Kuitunen N., Eklund L., Palotie A.,  
RA Pihlajaniemi T.;  
RT "Cloning of mouse type XV collagen sequences and mapping of the  
RT corresponding gene to 4B1-3. Comparison of mouse and human alpha 1  
RT (XV) collagen sequences indicates divergence in the number of small  
RT collagenous domains";  
RL Genomics 45:31-41(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=129/Sv;  
RX MEDLINE=20522048; PubMed=11068203;  
RA Eklund L., Muona A., Lietard J., Pihlajaniemi T.;  
RT "Structure of the mouse type XV collagen gene, Coll15a1, comparison  
RT with the human COL15A1 gene and functional analysis of the promoters  
RT of both genes";  
RL Matrix Biol. 19:489-500(2000).  
DR EMBL; AF261131; AAG27545.1; -.  
DR EMBL; AF261109; AAG27545.1; JOINED.  
DR EMBL; AF261110; AAG27545.1; JOINED.  
DR EMBL; AF261111; AAG27545.1; JOINED.  
DR EMBL; AF261112; AAG27545.1; JOINED.  
DR EMBL; AF261113; AAG27545.1; JOINED.  
DR EMBL; AF261114; AAG27545.1; JOINED.  
DR EMBL; AF261115; AAG27545.1; JOINED.  
DR EMBL; AF261116; AAG27545.1; JOINED.  
DR EMBL; AF261117; AAG27545.1; JOINED.  
DR EMBL; AF261118; AAG27545.1; JOINED.  
DR EMBL; AF261119; AAG27545.1; JOINED.  
DR EMBL; AF261120; AAG27545.1; JOINED.  
DR EMBL; AF261121; AAG27545.1; JOINED.  
DR EMBL; AF261122; AAG27545.1; JOINED.  
DR EMBL; AF261123; AAG27545.1; JOINED.  
DR EMBL; AF261124; AAG27545.1; JOINED.  
DR EMBL; AF261125; AAG27545.1; JOINED.  
DR EMBL; AF261126; AAG27545.1; JOINED.  
DR EMBL; AF261127; AAG27545.1; JOINED.  
DR EMBL; AF261128; AAG27545.1; JOINED.  
DR EMBL; AF261129; AAG27545.1; JOINED.  
DR EMBL; AF261130; AAG27545.1; JOINED.  
DR HSSP; P39061; 1KOE.  
DR InterPro; IPR000087; Collagen.  
DR InterPro; IPR001791; Laminin\_G.  
DR InterPro; IPR003129; TSPN.  
DR Pfam; PF01391; Collagen; 5.  
DR Pfam; PF02210; TSPN; 1.  
DR ProDom; PD000007; Collagen; 1.  
DR SMART; SM00282; LamG; 1.  
DR SMART; SM00210; TSPN; 1.  
KW Collagen.  
SQ SEQUENCE 1367 AA; 140454 MW; 40F259A3CD47C982 CRC64;  
Query Match 59.0%; Score 545; DB 11; Length 1367;  
Best Local Similarity 62.1%; Pred. No. 2.1e-45;  
Matches 105; Conservative 22; Mismatches 38; Indels 4; Gaps 1;  
Qy 7 QPVLHLVALNTPLSGGMRGIRGADFCQFQQAARAVGLSGTFRAPLSSRLQDLKSIYRRADR 66  
: | | | | | | | | | | : | | | | | | | | | | | | | | | | | | | : | | : |  
Db 1194 RPVLHLVALNTPVAGDIR----ADFQCFQQAARAAGLLSTFRAPLSSHLQDLSTVVRKAER 1249  
: | | | | | | | | | | : | | | | | | | | | | | | | | | | | | | : | | : |  
Qy 67 GSVPIVNLKDEVLSFSWDSLFRSGSQGLQPGARIEFSFGDRDVLHRHPAQKSVWHGSDPS 126  
: | | | | | : | | | | | : | | | | | : : : | | | | | | | | | | : | | : |  
Db 1250 FGLPIVNLKGQVLFNNWDSIFSGDGGQFNTHPIYSFGDRDVTDPSPQKVVVHGSNPFY 1309  
: | | | | | : | | | | | : | | | | | : : : | | | | | | | | | | : | | : |  
Qy 127 GRLMESYCETWRTTGTATGQASSLLSGRLLLEQKAASCHNSYIVLCIE 175  
| | | | | : | | | | | : | | | | | : : : | | | | | | | | | | : | | : |

Db 1310 GVRLLVDKXCEAWRTTDMVTFASFPLSTGKILDQKAYSCANRLIVLCIE 1358

Search completed: February 17, 2004, 09:57:53  
Job time : 38 secs